MASTER THESIS

SPEECH TECHNOLOGY TO SUPPORT INTERVENTIONAL RADIOLOGISTS BY ASSISTING THE HANDS-OCCUPIED PROBLEM

Ziyang Xia

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS & COMPUTER SCIENCE
DR. D.J.K HEYLEN

EXAMINATION COMMITTEE
Dr. Randy Klaassen
Dr. Khiet Truong

DOCUMENT NUMBER
HUMAN MEDIA INTERACTION -

18-08-2019

UNIVERSITY OF TWENTE.

UNIVERSITEIT TWENTE.

SPEECH TECHNOLOGY TO SUPPORT INTERVENTIONAL RADIOLOGISTS
– by assisting the hands-occupied problem

MASTER THESIS
Human Media Interaction
Faculty of Electrical Engineering, Mathematics & Computer Science

AUTHOR:

Ziyang Xia

CHIAR:

Dr. D.J.K Heylen

SUPERVISORS:

Dr. Randy Klaassen
Dr. Khiet Truong

ABSTRACT

The visual and tactile based traditional interaction modalities present challenges for the intuitive and independent manipulation of interventional radiologists. Currently, radiologists rely on an assistant to adjust devices when their hands were occupied with medical tools. This hands-occupied situation calls for the exploration of voice interaction solution to improve both working efficiency and user experience. User-centric design methods were adopted to identify user requirements and design opportunities. A study was designed to ascertain how challenging conditions influence work performance and user satisfaction in a simulated context. Results indicate high cognitive load when doing a precise placement under an indirect view. Potentials of cognitive overload by using voice commands to interact with the system were identified, especially when a random conversation occurs. In conclusion, we argue that a well-considered manner is vital when introducing voice interaction into interventional radiology. The cognitive state of radiologists and complexity of IR are two critical issues.

Keywords:
Interventional radiology, Voice Interaction, Hands-occupied problem, Cognitive load.

ACKNOWLEDGEMENT

I feel so proud and grateful when I recall the whole journey of the thesis work. It is a precious experience, so would like to express my sincere appreciation to those people who supported, understood, advised, and accompanied me.

I would first thank my supervisors Dr. Randy Klaassen and Dr. Khiet Truong for being supportive and helpful throughout the procedure. Thank you for reading my paper carefully and offering instructive guidance on explaining the problems and exploring the solutions. I would also like to thank my former first supervisor Dr. ir. H.J.A. op den Akker, who was also kind, patient, and give me a lot of encouragement at the beginning of the research. It was a great pleasure working under all of your supervision.

I'm very grateful to have all the radiologic experts from Yuncheng Hospital and designers in Siemens Healthineers for advising my work. Though I can't mention their names because of the confidential agreement, this research was not possible without your valuable inputs. I would also express appreciation to those voluntary participants for joining my testing and giving me useful feedback.

My special thanks go to my parents and sister. Thank you for your trust and encouragement as ever. It was you who help me to survive from those tough periods. I am thankful to all the lovely friends, colleagues, classmates, and crews from the university. Thanks for your kindness and patience for offering any kinds of support during the procedure. I would also thank my doctor, who was also vital to the accomplishment of the thesis.

Lastly, I would like to thank myself, thank you for always being insistent and brave. Because you never give up, now you made it!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

1 INTRODUCTION

1.1 Background

1.1.1 Introduction interventional radiology

Interventional Radiology (IR) is a precise target treatment that relies on a range of image-guided technologies to see the patients' anatomy[1]. The real-time x-ray technique fluoroscopy is one of the frequently used medical methods. Image-guided diagnostic and therapeutic procedures are at the core of this specialty [1], [2]. The principle of conducting IR is different from open surgery. In IR, doctors do not directly open the diseased situs, instead relying on an “indirect view” [3], [4]. Fluoroscopy can only scan part of the patient’s body in a fixed position. As a result, when radiologists need to adjust devices while manipulation medical tools, they struggle to allocate their hands [5]. According to a survey by the American Society of Radiologic Technologists (ASRT), radiologists are facing strong workload in clinic [6], [7]. Exploring alternative interaction modalities to benefit both working performance and user experience is thus significant [8].

1.1.2 Problem statement

One specific conflict drew our attention: how does an interventional radiologist adjust imaging devices when their hands are occupied with wires and catheters [9], [10]? We named this conflict the "hands-occupied problem" for consistency with the following texts. The solution in clinic is to assign an assistant to help the main radiologist adjust the related devices [11]. However, this solution not only limits the main radiologist's autonomy, but cost extra labor resources in hospitals where high workload in IR is already a problem [6], [12].

1.1.3 Possible solutions

Though the IR context presented multiple challenges for radiologists, many Human-computer interaction (HCI) experts started exploring solutions from different perspectives [10], [11]. Some prior research has already explored touchless solutions, such as gesture interaction [1], [13], robotic operating room assistance [14], telerobotic assistance [7], and laparoscopic voice control [15].

Voice interaction (VI), one imaginary technology from 20th century science fiction, has already been applied in different industries [16]. Automatic speech recognition (ASR) has experienced rapid growth both in software and hardware development [11]. The speed and accuracy of these speech recognition systems have significantly improved in the past decades. Users can interact with VI systems with less restrictions, larger vocabularies, more accurate speaker recognition, and more natural interactive methods [17]. For example, ASR technology improved

[1] https://www.bsir.org/patients/what-is-interventional-radiology/

report dictation in radiology departments [18], [19]; text to speech (TTS), a form of speech synthesis, is used in auditory instructions for car navigation [20]; and VI systems like intelligent voice assistants – including Amazon Echo, Apple Siri, and Google Assistant – are popular in the market [21].

However, speech technology is mainly used in IR for triggering or selecting functions while viewing images or during the setup procedure [22]–[24]. Some studies have conducted prior explorations of voice control for position tasks in laparoscopic and endoscopic procedures [15]. These studies suggested potential for applying VI to improve interventional radiologists' autonomy.

Most of the recent studies, however, lack consideration for the complexity of a real working context [7]. The hands-occupied problem remains unsolved. The user's severe cognitive load and complicated working procedure also require more user-centric investigations.

1.1.4 Focus of this research

The visual and tactile based traditional HCI modalities present challenges for the intuitive and independent manipulation of interventional radiologists. This situation calls for the exploration of alternative interaction modalities to improve both working efficiency and the user experience. The hands-occupied problem may be an opportunity to introduce VI into IR. VI technology, regarded as a human being's "third-hand," was usually proposed as a solution to a hands-busy situation. However, there are challenges to multiple aspects.

The first challenge is the variety of IR techniques. Different techniques require corresponding setups and experts [4]. Each working procedure contains various combinations of sub-tasks. Multiple devices and advanced technologies are applied in IR; it is a context of multi-modal interaction. Secondly, the team consists of several interdisciplinary experts according to different patient cases, so it is critical to define the commander. The communication among a team is also important. When there is a conversation, the system should distinguish between a voice command or just background noise.

The VI solution should be designed for the specific tasks and practical context. The new solution should not only help with the hands-occupied problem but also benefit interventional radiologists with a better working performance and user experience.

1.2 Research question

We argue that the exploration of new interaction modality is imperative. Voice interaction technology – which can replace hands through vocal commands – could be a potential solution. The complicated working environment and procedures still demand more user-centric investigations.

Therefore, in this research, we have conducted a study to answer this question: Can VI technology solve the hands-occupied problem in IR and further improve interventional radiologists' performance and experience?

1.3 Methodology

In this research, we used the User-Centric Design (UCD) methodology, a design process proposed by John Zimmerman et al. [25], as well as the empirical methods suggested by the book Research Method in HCI [26]. We followed the design progress in five stages. First, we uncovered fundamental background information, such as the medical principles, current challenges in IR, and solutions explored by recent studies. Second, we defined the target user, involved roles, context, and IR procedure with a context analysis. Third, we synthesized the results from the context analysis to define the problem and explicit the user requirements. We employed persona and scenario to analyze the design opportunities and challenges. Then, we developed a simulated testing setup to discover if the VI solution could improve the user's performance and experience in challenging conditions. The results of these tests were analyzed and concluded to deliver suggestions for future work.

We selected one or several UCD methods to achieve the aims of each stage. For example, we reviewed literature while conducting background research. When we investigated the detailed working procedure and user's condition, we combined information from literature, user interviews, questionnaires, and video records. Some methods – such as user persona, scenario, and interaction flow – were adopted to uncover the design opportunities and challenges. We also made use of some experimental methods, like control variable comparison tests. The following chapters introduce detailed methods for each stage.

1.4 Thesis structure

This thesis is presented in six chapters, including this first chapter that introduces the motivation, focus of research, and research question. This chapter provides readers with an overview of the work.

In Chapter 2, we summarize the background of IR and VI from the literature. Information about "what is IR" and "what are radiologists' problems," the current development of ASR, and the explored solutions are introduced to inspire our work.

In Chapter 3, we use interviews from related experts and reviewed videos to comprehensively understand the context. We present the results of the procedure, teamwork, and users' feedback of VI, respectively. Then, we discuss the results of the context analysis and literature review.

In Chapter 4, we define the hands-occupied problem in a more structured way and summarize user requirements for exploring solutions. The opportunities and challenges of introducing VI are also discussed.

In Chapter 5, a simulated setup demonstrates how challenging conditions influence a user's performance and experience when voice commands are offered. We introduce the design and procedure of the study, then analyze the data and conclude accordingly.

In Chapter 6, we summarize the work completed and discuss the results and limitations. We also propose several directions for future studies.

2 LITERATURE REVIEW

In this chapter, we introduce the existing results on three topics from the literature, IR (Section 2.1), ASR technology (Section 2.2), and VI solutions in minimally invasive treatments (Section 2.3). The first topic – IR – is introduced to answer three questions: “What is IR?”; “How does it work?”; and “What are the current problems from radiologists’ perspectives?”. We then summarize the fundamentals and current development of ASR. Prior explorations of VI in IR and similar techniques are demonstrated afterward. In the end, we summarize the recent studies to denote any found and unsolved issues (Section 2.4). We propose our research focus based on these recent studies.

2.1 Interventional radiology

2.1.1 Overview of interventional radiology

Interventional radiology is a precise target treatment that relies on a range of image-guided technology to view the patients’ anatomy[2]. Previously, the radiology department only used image techniques to deliver exact diagnoses to referring doctors [27]. It was regarded as a supporting department because it did not run treatments. Since 1964, however, a new medical technique with a smaller incision and lower risk has been gradually applied to multiple diseases [28]. According to the book PRINCIPLES AND PRACTICE OF CLINICAL RESEARCH (THIRD EDITION) [29], the modern medical definition of IR is:

> *“A therapeutic and diagnostic specialty comprises a wide range of minimally invasive imaging-guided therapeutic procedures, as well as invasive diagnostic imaging.”*

The image-guided technology introduced by the British Society of Interventional Radiology (BSIR) is comprised of real-time x-ray, ultrasound, computed tomography (CT), or magnetic resonance imaging (MRI)[3]. Fluoroscopy is the technique relying on real-time x-ray [3], [30]. Currently, IR is a set of new techniques charged by the radiology department in hospitals. Professionals mastering IR techniques are interventional radiologists [30].

These IR techniques emphasize the features of relying on image-guided techniques and precisely targeted treatment [28]. Conducting IR is different than open surgery. Radiologists interpret a patient’s anatomy via an “indirect view” instead of opening a diseased situs [31]. They then precisely repair a variety of disorders by manipulating medical tools to reach the site of the problem [32].

[2] https://www.bsir.org/patients/what-is-interventional-radiology/

[3] https://www.radiologyinfo.org/en/info.cfm?pg=professions-diagnostic-radiology

Usually, both their hands are occupied with holding guiding wires and catheters [33].

Interventional radiology provides vital remedies to patients. General anesthesia is optional in most IR cases. Thus, IR has lowered risks and reduced morbidity. Compared with open surgery, the other advantages of IR include shorter hospital stays, lower costs, greater comfort, and quicker convalescence[4]. The work volume of IR has increased in recent years [12]. These IR departments are facing an understaffing problem [6], [37], [38]. According to the radiologic sciences staffing and workplace survey by ASRT in 2017, approximate 2.62 patients were imaged every day in cardiovascular intervention (CVIT) [35].

New techniques have been increasingly developed, since Charles Dotter first used IR to dilate a stenosis in 1964 [28]. Interventional radiology has successfully cured diseases in the vascular, pulmonary, biliary, gastrointestinal, genitourinary, and central nervous systems. Notably, the range of conditions covered by IR is continuously expanding [19]. The boundaries between IR and surgery are mixing [36]. Furthermore, IR procedures have embraced increasingly minimally invasive therapies. Normal interventional procedures consist of several medical options including diagnosis, drug treatment, and open surgery [28], [37]. Each IR technique requires specific expertise and arrangement [38]. A restricted research scope is necessary.

We chose Percutaneous Coronary Intervention (PCI) [39] techniques as our focus, because PCI covers a complete IR procedure from disease diagnosis and interpretation to treatment. As another reason, blood vessel diseases are one of the most common and typical diseases treated by IR. Angiography, balloon angioplasty, and stents are three techniques in PCI. Angiography is a typical diagnostic technique for blood vessel diseases. Balloon angioplasty (sometimes called balloon) is a common, minimally invasive procedure to open up the blocked or narrowed arteries and restore blood flow. When a balloon fails, stents are typically used.

The mixed usage of interventional techniques and open surgeries demands a multi-modal interactive environment for an interdisciplinary team, called the hybrid operating room. Figure 1 shows one example of the facilities in a hybrid operating room. A fusion of multiple facilities can be identified [38].

[4] https://www.radiologyinfo.org/en/info.cfm?pg=professions-diagnostic-radiology

Figure 1 : An example of IR operating room with facilities like a patient table, a fluoroscopy, a moveable ultrasound, a contrast injector, several displays, and other related devices[5].

A hybrid operating room often consists of an operating room and control room [11] (see Figure 2). In the operating room, the patient table is positioned in the center, providing the pivot point for the room, surrounded by various medical equipment. The fluoroscopy is commonly located near the patient table. It is also called a C-arm, as it looks like a letter “C”. Above the table, there is a ceiling-mounted display consisting of four to six screens. Displays are moveable so all team members can have an unobstructed view [11], [19], [40]. Other equipment – such as the instrument trolley, foot pedal, contrast dye injection pump, ultrasound machine, anesthesia equipment and mobile radiation protection screen – are placed nearby. In the corner, there are scrub sinks and paraphernalia shelves with different kinds of wires, catheters, puncture needles, and other medical instruments.

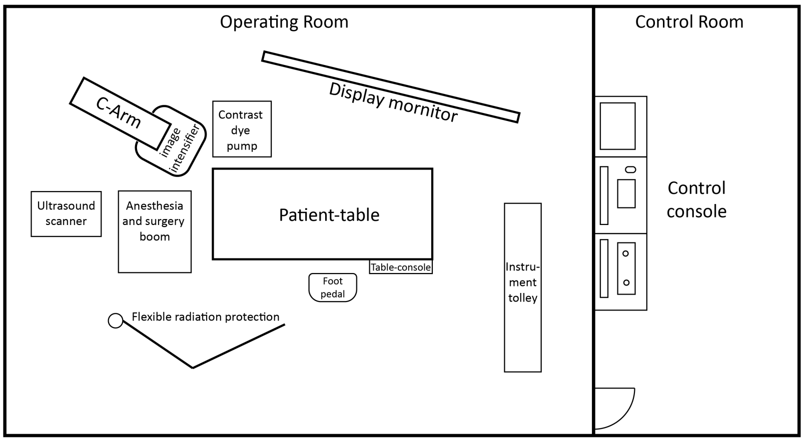

Figure 2: A layout of a hybrid operating room [11], [19], [40].

Doctors in the control room support the IR procedure via a set of computers and control consoles [38]. As shown in Figure 3, the operating room and control room are usually located next to each other with a transparent glass window to

[5] https://www.uniklinik-ulm.de/unfall-hand-plastische-und-wiederherstellungschirurgie/verletzungen-im-kindesalter.html

separate them. Operating rooms should be sterile during the operation, while control rooms do not have this requirement.

Figure 3: An example of control room equipped with several computers and control consoles [6].

These facilities were used according to specific medical needs during IR [37]. To clarify the fundamental working principle of related facilities and image technology, we present six typical medical actions from the literature in the following paragraphs. There is no certain order to these actions. These actions explain how different technologies function together to make a success of treatment.

1. Seeing patients' anatomy via fluoroscopy

The fluoroscopic imaging provides continuous x-ray images of the patient's anatomy. The continuous images are displayed on a screen in front of the radiologists (see Figure 4). However, blood vessels are not visible under the fluoroscopy. X-ray images are only 2D, lacking a 3D image of the vessel structure [33], [41]. For limited radiation exposure, fluoroscopy is both low dose and low resolution [42]. It can only scan part of the patient's body in a fixed position. Therefore, radiologists have adopted several methods to address these limitations.

Figure 4 : A radiologist is adjusting the joystick with his eye staring at the screen, another doctor is manipulating other device and look at the screen as well[7].

[6] https://www.siemens-healthineers.com/nl/angio/options-and-upgrades/components-and-options/artis-cockpit

2. Injecting contrast dye to make blood vessels visible

Contrast dye is used via a catheter along with a guide wire to highlight the blood vessels [41]. The wire can be seen under the fluoroscopy, while the contrast dye visualized the blood flow for identifying any stenosis or blockage [43]. The contrast medium can be injected either by hand or using a pump[8]. Actual contrast volume depends on the patient's condition, the catheter size, and the speed of injection. Therefore, the contrast dose should be individualized for each case [44]. These parameters are usually preset during the setup period [37].

3. Adjusting a table or C-arm to find the optimal positions

Radiologists must find the optimal position and angle to “see” the diseased point, as fluoroscopy can only show a 2D image of a certain part of the body. Therefore, flexible, moveable position system is critical. The position system contains two components: the moveable C-arm and patient table. Both can be adjusted manually forward, backward, left, right, up, and down to allow fast and precise movements during IR [8]. Figure 5 demonstrates how radiologists use the joysticks located in the table console to adjust its position. However, as the scanner can only scan part of the patient’s body in a fixed position, radiologists must move the fluoroscopy and table accordingly throughout the procedure.

Figure 5: The table console in Siemens Artis phone[9].

4. Using the foot pedal[10] to capture high-resolution images

Radiologists use a foot pedal to capture high-resolution x-ray images called angiograms to assess blood flow, or detect any blockages. They capture a sequence of angiograms to show the blood flow when the contrast dye flows through the arteries [9]. However, radiologists cannot always take high-resolution images because they must decrease radiation exposure to prevent in severe occupational disease [42]. There is a trade-off between the number and quality of

[7] https://www.siemens-healthineers.com/refurbished-systems-medical-imaging-and-therapy/ecoline-refurbished-systems/angiography-ecoline/artis-zeego-eco

[8] https://www.radiologyinfo.org/en/info.cfm?pg=angiocerebral#how-its-performed

[9] siemens.com/artis-pheno.

[10] https://www.transamericanmedical.com/c-arm-wireless-hand--foot-switch.htm

the images and the amount of radiation exposure. High-resolution images are executed only when absolutely necessary. When it is the time to capture images, radiologists have their hands full holding and manipulating the wires, and their eyes are fixed firmly on the fluoroscopy screen watching the contrast dye move. While watching, they combine what they see on-screen with what they feel in their hands. Timing the trigger is a delicate procedure**.**

Figure 6: The foot pedal switch[11].

5. Viewing the sequence of angiographies to see the blood flow

Radiologists inspect particular features on a single image as well as the nature of the blood flow shown by sequenced images. Moving through the images corresponds to moving through time. A sequence of images consists of the progression of contrast dye flow through the arteries. For example, a large section of artery has not filled or blood is flowing into minor branching arteries; this pattern indicates areas of blockages or narrowing. Radiologists interpret the pathology not only through these visuals but also by the feeling as they proceed through the images. Mapping blood flow with the speed and direction of viewing images is personal and delicate. It can be done with a mouse at the computer or by control in the control room. Radiologists sometimes go to the control room to view the images [11].

Figure 7: In the control room that technologists are selecting images for radiologists[12].

[11] Siemens.com/ artis zeego ego.

[12] https://www.youtube.com/watch?v=w0EQQzIwk-E.

6. Using digital subtraction angiographies to dismiss the inferring tissues

Radiologists sometimes add masks or ask for the masked angiographies to guide the insertion during the treatment. Digital subtraction angiography (DSA) is a well-established 2D imaging technique visualizing of blood vessels in the human body [44]. As shown in Figure 8, background structures such as bones are largely subtracted by a mask image from the original images. The mask images are usually images acquired prior to injection. Often, DSA is combined with road mapping. The frame with maximum vessel opacification in a DSA sequence is identified, which becomes the roadmap. The roadmap is useful for the precise placement in complex and small vasculature. Fluoroscopy alone may not adequately visualize small wires and vessels in the distracting underlying tissue. Radiologists sometimes go to the control room to make the edition on their own [11]. They try to outline the diseased spots as clear as possible by manually move the position and brightness of roadmap overlay [9].

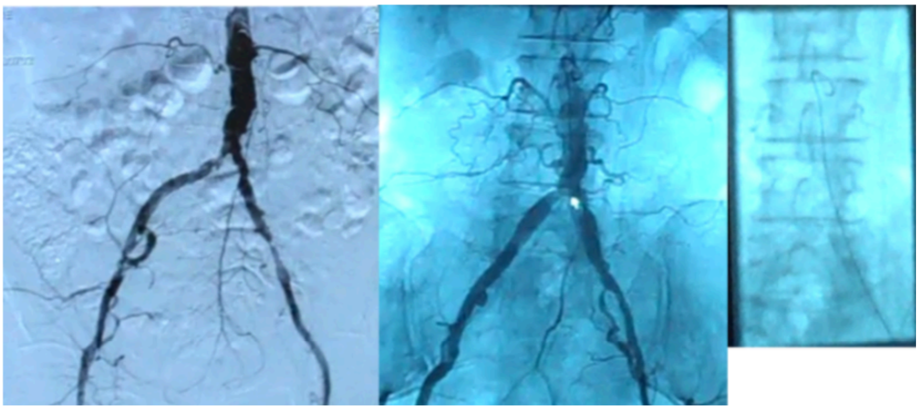

Figure 8: An example of DSA images, the left is the edit result, it was made from the original image (in the middle) subtracting the pre-took mask image (the right one) [11].

In brief, we noticed a close interrelated logic among a sequence of actions. In a hybrid operating room, multiple facilities function together. Team cooperation is critical to make a successful treatment. The two hands of radiologists were needed to controlling multiple instruments in different actions. However, the details of hands allocation were unclear.

2.1.2 Related work

I. Problems in IR

Several researchers have conducted field research to either improve workflow or explore the potential of touchless solutions. These studies reminded us of the complexity of IR and inspired us with potential solutions. We present a learning curve of understanding our research question when introducing the related studies.

For example, Hübler et al. [10] observed 25 neuroradiological CT interventions to identify the actions frequently used and find sequences of actions that often occur. They distinguished 99 different actions (buttons, functions, and verbal commands). Quantitative analysis revealed that the most frequently used functionalities are related to table position control, C-arm position control, and aperture control. The most frequent interactions and their related actions indicate a closely interrelated logic between a sequence of actions handled by different roles. Close teamwork among the IR team, especially the physicians and the technician, is thus critical.

Hübler et al. [10] also amassed the following problems when discussing findings with three physicians and one technician: First, the user tried to interact with an interaction device out of reach or blocked by another user. Moreover, users also had to leave the sterile area to get additional patient data from a workstation in the control room. Additionally, the most frequently used interaction elements should be easy to find. They were mostly displayed on the start screen. As they noted, functions that cannot be found quickly enough – e.g., in critical situations – are not used during an intervention, even if they might be helpful.

We also found further research answering several of the questions above. R. Johnson et al. [11] observed six IR procedures over 2 days, focusing on work practices and interactions in the angiography suite of a large British hospital. This research explored the potential for touchless interaction in IR through a study. Johnson explains why radiologists leave the sterile operating room. Radiologists prefer to use a mouse to view captured images or edit DSA images. These two tasks are also introduced in the six medical actions (see Section 2.1.1). Radiologists prefer independent manipulation because viewing the sequence of angiograms is combined with the rhythm of clicking a mouse. Additionally, editing DSA images is a dynamic procedure. Radiologists must realign the position of the mask slightly in case of mis-subtraction. This procedure also helps radiologists identify disease, as certain features can only be highlighted by repeatedly revealing or removing the masks. Manual adjustment requires leaving sterile area, but it supports the medical interpretation. From above, we find that an assistant cannot replace a radiologist for all interactions. Moreover, radiologists' requirement of independency is due to more specific purposes.

Johnson [11] elaborates upon the position task, where radiologists' hands are often typically occupied by holding medical tools. The moment radiologists insert the guide wires and catheters, they simultaneously need the corresponding position of the patient's anatomy – the hands-occupied problem of our research. Johnson found that communication between the radiologists and assistants is not always fluent and easy. They also provided a vivid conversational example when instructing the assistants in finding the proper position:

> *"Radiologist: North please... thank you stop there. Wire out, I'll have a J wire back please, thank you. Open sides a little bit for me now please. And up down a bit more. Thank you, keep going up down. Thank you... Fine, wire out thank you. P you're going to*

> *get a bit of that warm feeling from that x- ray dye I told you about. OK, if you feel like you're going to wee don't worry.*
>
> *Patient(P): I felt like it earlier on.*
>
> *Radiologist: Did you? Well haven't put any in yet, but I'm going to now. So... Yep, did you get that? OK it will be a bit more intense when we get the proper pictures alright? OK.*
>
> *Patient: It feels as though it's molten actually.*
>
> *Radiologist: Yes, it will. Absolutely, it will do a little bit. That's great, OK let's get set up for the peri run now then. Better bring the table vertically up M [assistant]. Do you want to go vertically up then I can get the table up a bit higher? That's good we can keep it this way round cos then we can get that all on. OK M [assistant] that's fine. Just got to get some preliminary set up pictures sorted now P. Those'll be ready to go. Err sorry, hang on hang on. Just go back up, the left side is the one we're more interested in, and it mustn't come across. That's good, sorry. Yeah Ok?*
>
> *Assistant: Sorry, I took your...*
>
> *Radiologist: I know I take my foot off. Can we get the legs a bit closer together? P, we just need to try and get your knees a bit closer if we can."*

In this example, "North" indicates movement toward the patient's head; "south" denotes movement toward the patient's feet. The main radiologist converses with the patient and the assistants. The exchanged information includes how to adjust the intensifier and the expected direction, the timing of the start/stop point, the purpose of adjustment, etc. The radiologist provides random but specific instructions when communicating; these parameters are dependent on the procedure and factors, such as the size of the patient.

The aforementioned study finally provides two opportunities for touchless control; one is image-based interactions behind the radiation screen, while another is the monitors above the x-ray table. The paper did not propose a specific solution.

Two other studies focused on severe cognitive load in IR, attempting to discover solutions. Anderson et al. [33] addressed the difficulties of controlling instruments under an indirect view and introduced a system to train interventional radiologists. The indirect view features require significant practice in clinical settings, as stents or balloons are accessed by long instruments. The indirect view results in greater physical and mental consumption than open surgery, as internal structures are usually visualized through a two-dimensional optical system, diminishing depth perception. Moreover, the tissues being operated on are accessed by long instruments, which results in decreased tactile sensation and the amplification of any tremor. Furthermore, there is a limited range of motion for the radiologist's instruments because of the Fulcrum effect (bias in the perceived stiffness, especially with novices) [45].

Radiologists split their attention between the situs, monitor, and controls. Herrlich et al. [46] focused on the cognitive load of this split attention between the navigation system and the situs. It is a typical multi-modal interactive context; interventional radiologists rely on such multi-modal interaction with diverse medical equipment. Internal structures are usually visualized through a 2D optical system that diminishes depth perception. The physical and mental consumption of interpreting these 2D images should be considered.

To summarize, several problems were collected from related studies.

1. Users have to leave the sterile area to retrieve additional patient data. The reasons are explicated by Johnson et al. [11]. This situation uncovers the radiologists' desire of independent manipulation. Radiologists sometimes prefer manually adjustment on their own rather than relying on an assistant, is due to specific purposes (such as, to map blood flow, to outline diseased spots).
2. Functions that cannot be found quickly enough are not used during the intervention, even if they are useful. The reason is still unknow. Cognitive overload might be the reason.
3. Recent studies denoted severe cognitive load on the radiologists. The in-direct view and split attention require extra cognitive resource when radiologists making a treatment.
4. Radiologists rely on an assistant to adjust related devices when their hands are occupied. Close teamwork among the IR team is critical because of the interrelated logic of medical actions.
5. Assistants could not replace radiologists for all interactions. The communication between the radiologists and assistants was not always fluent and easy.
6. Lastly, radiologists try to interact with an interaction device out of reach or blocked by another user. The reason is still unknown. We wonder if radiologists are limited with spatial mobility.

These problems in literature are not directly linked to hands-occupied situation, but their root cause may be critical to our research. The root reasons remind us the hands-occupied problem is not a single issue that can be solved by simply adding a "third-hand". Actually, the hands-occupied problem was complicated and resulted from multiple factors. A considerable fusion of VI solution and IR context is vital.

II. Interaction alternatives for interventional radiology

Though the current IR context presents multiple challenges for radiologists, many HCI experts begin exploring solutions from different perspectives to assist the progress. Tan et al. [1] and Lannessi et al. [13] tested a prototype controlled by gestures that focused on the image-review procedure as well by calling for prior and real-time patient image. However, the results did not obviously favor the new solution, as the majority of the participants regarded the solution as moderately difficult.

Mewes et al. [7] systematically reviewed of all recent work on touchless solutions in IR before 2016. Summarily, most of the authors described methods for the touchless manipulation of medical image data. Other objectives were laparoscopic assistance, telerobotic assistance, operating room control, robotic operating room assistance [14], and intraoperative registration. The technologies used included the Microsoft Kinect 1 structured-light-based range camera, the most popular device for touchless intraoperative gesture control; stereo cameras; Leap Motion controller; body-worn inertial sensors; a camera or webcam; voice recognition; eye tracking; the Intel Real-sense Creative structured-light camera; and a time-of-flight range camera. However, most (92% of the reviewed papers) were not tested in a real clinical environment. These studies noted the demands on the further development of hands-free interaction with better fusion as well as the reduction of interaction steps.

Specific related studies about speech technology in IR or its similar medical techniques (see Section 2.3) are introduced after the overview of ASR (see Section 2.2).

2.2 Automatic speech recognition

The IR procedure is delicate, and multiple interactions are combined. As the focus of our research is to explore a VI solution, the background of VI is imperative.

2.2.1 Overview of speech technology

The notion of ASR is also known as speech recognition or computer speech recognition, which is converting spoken language to text and performing any required task [47], [48]. VI, including voice recognition, stresses the modalities of how humans interact with the target devices, while ASR provides technological support for the success of VI [49]. The systems are composed of microphones that convert sound into electrical signals, sound cards that digitize the electrical signals, and speech engine software that converts the spoken language into text [50]. Voice interaction can be divided into three uses: commanding the computer; entering information; and communicating with other people [49].

In our context, the current interactive environment is already a multi-modal interactive environment. Using VI to help with the hands-occupied problem is actually introducing speech and auditory modality as voice control into the current multi-modal interactive system. Modality fusion is often the center of a multi-modal system [51]. Here, input comes from the keyboard and mouse in the control room as well as the touchscreen and manual gestures of the control levers and foot pedal equipped on the patient table. The output is mainly visual modality via the large monitors, with tactile feedback from the physical buttons and controls. Therefore, the current development of ASR and the design principle of multi-modal interactive systems are both important.

2.2.2 Related work

I. Features that influence the capability of recognition.

The elements that influence recognition speed and accuracy are speaker dependence, and vocabulary size, continuity of speech, intuitiveness, and simultaneity [16], [18], [47], [49].

- Speaker dependence

Systems that have historically met the highest recognition accuracy are speaker-dependent systems, wherein each user must create an individual template that reflects the way he or she speaks. Speaker-independent systems use generic models to recognize speech from any user; the recognizer works for anyone's voice with specified vocabularies [14]. This approach is advantageous in that it does not require individual operators to train the system to recognize their voices. However, because the templates are not user-specific, accuracy rates are usually lower [52].

- Vocabulary size

The vocabulary size influences the recognition speed and accuracy of a voice interaction system as well. Generally, a small vocabulary contains tens of words, a medium vocabulary contains hundreds of words, and a large vocabulary includes thousands of words. The current vocabularies of various ASR systems range 20-40,000 words [47]. Another factor related to vocabulary size is the complexity of the grammar or words in the vocabulary. A vocabulary with little variation in grammar or a rigid system of word sequencing are less complex than a vocabulary with a wide variation in grammar and little predefined structure to word patterns or sequences. However, too many grammar rules can reduce the naturalness of communication [47]. If the user's input is restricted to the recognizer's vocabulary and style limitations, recognition performance in the latter applications is higher [17].

- Continuity of speech

A continuous speech system is more natural to use as words do not need to be separated by pauses. As a result, however, continuous speech is more difficult for computers to recognize accurately. A small vocabulary and tightly constrained grammar are especially important when a system is intended to support continuous speech [52]. Another critical issue in continuous speech is word coarticulation, where the beginning and ending phonemes (the individual sounds of a word) of a word are affected by the surrounding words. Because of these problems, continuous speech systems have historically had a lower accuracy rate than discrete speech systems [52]. Isolated word systems require a deliberate pause between each word, but pausing for 0.1 second after each word is unnatural and tiring. However, accuracy rates are generally better with isolated word systems than systems using continuous speech. Isolated systems work best with vocabularies that consist largely of individual command words.

- Intuitiveness (spontaneous)

Most speech-centric multi-modal interaction systems are user initiative (for example, through push to talk) [16]. A user-initiative system can be modeled as a set of asynchronous event handlers. In a more advanced system, the system should also actively interact with the user to ask for missing information (mixed initiative). Multi-modal dialog systems should be developed within a statistical framework that permits probabilistic reasoning about the task, context, and typical user intentions [53].

- Simultaneity

Some systems are incapable of “listening” to speech input while simultaneously producing speech. With these systems, it is impossible to interrupt a prompt, and the user must wait until the prompt finishes before responding. Systems that detect speech during prompts and truncate the prompt accordingly provide more natural interfaces, but adding this capability may also lead to increased ambiguity surrounding. As a result, the user interface for a system with prompt talk-through may require a more complex strategy for determining whether the interrupting response is a delayed response to prior prompts or an anticipatory response to the interrupted prompt [17].

Generally, there is a trade-off between natural interaction and recognition speed and accuracy. In our context, balancing the reduced cognitive load of ASR use and ensuring the performance of the system is critical.

II. System capabilities and limitations

Advances in computer hardware and software have improved the ASR systems used for radiology reporting [16], [54]–[56]. Voice dictation in radiology reporting is a use case for entering information, providing knowledge of the input capacity of current ASR systems [50], [52], [57]. Prototypes of speaker-independent real-time ASR have demonstrated the ability to understand naturally spoken utterances with vocabularies of 1000–2000 words and larger [16]. This technology also has a significant impact on the ability of healthcare providers to operate more effectively and provide a better level of patient care [50].

The other speech applications demonstrate a wide range of voice input capabilities as different features are adopted [58], [59], [60]. The current success still requires using a small to medium vocabulary – less than 1000 words – and well-defined grammatical rules. Research has shown that prototype spoken-language-understanding systems with vocabularies of 300–1000 words can accept spontaneous continuous speech with user enrolment [17]. However, this capability was restriction to limited domains, like traveling plans or directional guidance. The current understanding rate is around 70%. Dictation systems with large vocabularies work best when the operator uses only a subset of keywords [47]. To attain a sufficiently high level of recognition accuracy in field tests, spoken input has been severely constrained to allow only a small number of possible words at any given time.

III. Cognitive effect of VI

We collected researches about cognitive load in multi-modal interaction context. According to Baddeley's working memory model [61], working memory has separate stores (perception channels) for visual and auditory information, and each store has a limited capacity. Therefore, the capacity of working memory can be better used when both channels perceive information. This theory is known as the dual-channel theory. Cognitive load occurs when someone needs to process information in their working memory – for example, during a task, test, or instruction. As working memory is limited, a high cognitive load can occur. Furthermore, schema acquisition and automation are hindered when someone devotes too much mental effort to a learning activity. When the load is too high, it may result in cognitive overload [62], [63].

Thus, voice-centric systems could improve performances by using both perception channels [59], [20]. Examples like well-designed speech-based in-vehicle systems allow drivers to keep their eyes on the road and their hands on the wheel, making it possible to improve driving performance and reduce accident hazards [20]. A study by Sistrom [64] noted that radiologists can recapture efficiency and cognitive focus by dictating while viewing images without the ''look away'' problem inherent in other interfaces.

However, voice-centric systems could also result in cognitive overload when too many channels are provided and tasks are too complex. Cao et al. [65] have already demonstrated that multi-modal presentations should be created in a cognitively aware manner, especially in a high-load HCI situation where the user's task challenges the full capacity of human cognition. In our situation, radiologists' cognitive resource is distributed to various tasks. The mental consumption of users' must be considered when exploring VI solution.

2.3 Speech technology in minimally invasive treatment

The growth of image-guided procedures has led to an increased need to interact with digital images and related devices [7].

I. Voice commands for image viewer in IR

Prior studies of VI have explored IR. In the area of medical image viewers, VI is largely explored as independent commands or combined with other interactive modalities. In this sub-task, radiologists may go to the control room to view the captured images. However, while the hands of radiologist are not occupied, they must be kept sterile. Some touchless solutions – including voice – would help in avoiding the use of mouse and keyboard.

Schwarz, Bigdelou, & Navab [22] introduced a technique that collects pose data from multiple body-worn inertial sensors and classifies them as low-dimensional body gestures. This system was later extended with a voice-based and handheld switch unlock method [23] Using a handheld switch to unlock the interaction is preferred over a voice trigger, possibly due to a faster response time. Hötker et al. [66] drew another comparison. Six voice commands and six hand gestures with the same functions for medical image manipulation were implemented. A user study with 10 subjects indicated that voice commands (97%) were better recognized than the body gestures (88%) with an overall false-positive rate of 30%.

Apart from these papers, we also found that Wipfli et al. [67] compared three possible different interaction modes for the sub-task of image manipulation in IR: a gesture-controlled approach using Kinect, oral instructions to a third party dedicated to manipulating the images; and direct manipulation using a mouse. The result illustrated that both efficiency and satisfaction were greatest when using the mouse, followed by gesture control and oral instructions. But no significant difference in effectiveness was found. This research concluded that when the mouse cannot be used directly during surgery, gesture-controlled approaches appear superior to oral instructions for image manipulation.

Meanwhile, Ebert et al. [68] introduced a different medical image viewer control. Voice commands and range camera input were mapped onto keyboard and mouse events. With vocal commands, the interaction modes could be switched. The images were then manipulated (windowing, scrolling, and moving) with arm gestures. A comparison with mouse interaction in a user study of 10 subjects resulted in mouse interaction being 1.4 times faster than gesture interaction. An overall usability rate stated 3.4 out of 5, with 3.4 for accuracy of the gesture control and 3 for the accuracy of the voice control.

II. Voice control of laparoscopic and endoscopic devices

Another recent study occurred in the field of endoscopy. Endoscopy is a similar technique that allows the surgeon to see the patient's anatomy with the assistance of a video camera through a small incision [69]. Both IR techniques and endoscopy are guided by and rely on real-time captured images to see internal anatomy. However, the way they see internal anatomy is different. Endoscopic surgeons view the anatomy through a miniature digital video camera placed at the end of a laparoscope inside the body, while radiologists see it though real-time fluoroscopy outside the body. But both techniques have a common problem – the need to control these devices when hands are otherwise occupied.

One approach that has already been explored in endoscopy is AESOP [13] (automated, voice-activated endoscopic system for optimal positioning) [15]. This system controls the video camera inside the patient according to voice commands provided by the surgeon. The AESOP system belongs to a robotic device called ZEUS, the first robotic device to assist surgeons in the operating room. It has three robotic arms controlled remotely by the surgeon. One is AESOP; the other two acts like extensions of the surgeon's arms, following the surgeon's movements while allowing for more precise execution by scaling down movements and eliminating any tremors resulting from fatigue. The success of this system includes allowing the surgeon to perform the operation seated in an ergonomic position, eliminating fatigue and frustration from leaning over the patient in an awkward posture for hours.

El-Shallaly et al. [70] compared routine laparoscopic intervention with a commercial ASR interface. Via voice commands, the light could be activated, the camera established and white balanced, and the insufflator controlled. After treating 100 patients with and without the system, the authors concluded that significantly less time was spent switching components on and off, compared to manual control.

Salama & Schwaitzberg [71] investigated the availability of a voice-activated control for light, camera, and insufflator in comparison with an assistant. As a result, while the nurse was not immediately prepared to execute commands in 77% of occurrences, voice commands were given to the voice control system. Thus, voice commands appear to make laparoscopic interventions more productive.

III. Voice as operating room control and logging

One exploration addressed voice controllers for image-guided surgery and personalized interactive visualization. Beristain et al. [72] introduced a system that permits planning the display preferences. It can configure different monitors and assign input sources. Further research by Alapetite focused on the two hands issue when placing anesthesia [73]. A voice-recognition-based anesthesia record system is developed. The doctors must use both hands to ascertain the patient's health. The system allowing anesthetists to control liquid flow and log events with spoken commands from a fixed dictionary. Command recognition starts after an unlock keyword. It automatically stops after a fixed period. Free text input is also possible. In the evaluation, conventional keyboard and touchscreen input were compared to VI. The study demonstrated that mental workload is decreased by natural language interaction.

[13] https://spinoff.nasa.gov/spinoff2000/hm1.htm

2.4 Conclusion

In this chapter, to understand our research background, we introduced the fundamentals of IR, the development of VI, and prior research into VI in minimally invasive treatment procedure. We also explored literature that described the IR working principle, the problems of IR presently, and the offered solution. First, we identify several IR features relating to the hands-occupied problem.

2.4.1 Features of interventional radiology

Two typical features of IR are reliance on image-guided techniques and precisely targeted treatments. Radiologists interpret the patient's anatomy via an "in-direct view" powered by the fluoroscopy. Radiologists acquire the optimal view accordingly by adjusting the imaging devices throughout the procedure. Then, radiologists can then precisely repair a variety of medical disorders by manipulating the wires and catheters to reach the site of the problem. Both of their hands are usually occupied in doing so.

Besides, the closely interrelated logic among a sequence of actions is another feature we must emphasize. In a hybrid operating room, multiple facilities function together by a close cooperated team. The two hands of radiologists were allocated to controlling multiple instruments in different actions. Positioning the related devices is one of the moments when radiologists' two hands are occupied.

The fourth feature is, radiologists rely on assistants to control devices when their hands are occupied. We identified a continuous communicating process between the radiologist and the assistant in literature [11]. Radiologists' desire of independent manipulation is uncovered when they leave the sterile area to view or edit images.

However, the details of hands allocation were unclear. We need to uncover more details. Currently, IR techniques are used to treat multiple diseases. Each IR technique requires specific expertise and formats. To limit the research scope, we chose one target technique named PCI.

2.4.2 Problems identified in literature

IR techniques do provide multiple benefits to modern medicine. However, they also raise challenges for the radiologists who are responsible for such a complicated technique. Four summarized problems were collected from the relevant studies (see the list below). These problems remind us the hands-occupied problem is complicated and results from multiple factors. A considerable fusion of VI solution and existing interaction modalities in IR context is vital.

1. Radiologists have to leave the sterile area to retrieve additional patient data. They prefer view and edit images independently, because it can help map the blood flow or outline diseased spots.
2. Functions that cannot be found quickly enough are not used during the intervention, even if they are useful. Cognitive overload might be the

reason. The in-direct view and split attention require extra cognitive resource when radiologists making a treatment.

3. Radiologists rely on an assistant to adjust related devices when their hands are occupied. This situation is our focus as well. Close teamwork among the IR team is critical because of the interrelated logic of medical actions. Assistants could not replace radiologists for all interactions. The communication between the radiologists and assistants was not always fluent and easy.
4. Radiologists sometimes try to interact with an interaction device out of reach or blocked by another user.

2.4.3 Technology discussion

The current IR interactive environment is already a multi-modal interactive one. Using VI for the hands-occupied problem is introducing speech and auditory modality into the current multi-modal interaction system. Research demonstrates that speech technology has already experienced substantial growth over the past years. Some features of VI can be designed to improve recognition capability. However, there is a trade-off between natural interaction and recognition speed and accuracy. An effective cognitive-aware manner is suggested, especially in a high-load multi-modal interactive context. These examples of VI have shown conflicting results for increasing or decreasing cognitive load. Success can also be found in the precise control room. Problems may potentially be solved by well-designed voice-involved multi-modal interactive system. Balancing the reduced cognitive load of ASR use and ensuring the performance of the system is critical in our research.

Several researchers have already explored touchless solutions in IR. Thereinto, speech technology is mainly used for triggering or selecting functions during image viewing or setup procedure. Though Mentis et al. [24] proposed voice commands for trigger functions, and hand or body gestures for the continuous manipulation of parameters. Prior work has explored VI for positioning tasks in laparoscopic and endoscopic devices. Other success could be identified in configuring monitors, controlling liquid flow, and logging events during anesthesia. These explorations displayed the potential of VI in complicated tasks.

Last but not least, many of the recent studies still lack consideration for the complexity of a real working context [7]. The complex working environment and procedure also demand more user-centric investigations. The reasons and details of the hands-occupied situation are still unclear; the users' cognitive state and the team work require more consideration. Next chapter, we conduct a context analysis to understand IR context and the hands-occupied problem more meticulously and precisely.

3 CONTEXT ANALYSIS

In the previous chapter, we discovered that the complexity of IR was not thoroughly considered. Some specific issues – such as team cooperation, state of the target users, reasons for problems, and the working context – were still unclear. In this chapter, according to UCD theory, we introduce a contextual analysis procedure to uncover these answers. We utilized two user research methods: interviewing related experts and observing video records. During the interviews (Section 3.1), open questions and questionnaires were employed accordingly to reveal results across different topics. The detailed results are thus summarized (Section 3.1.5). However, as we were limited in observing a real treatment, we searched online video records to reconcile the missing details (Section 3.2). We then discuss the results of the context analysis and literature review (Section 3.3), and we use the integrated results as a foundation for the design analysis in the next chapter.

3.1 Interview

3.1.1 Motivation

The interview was the primary research method in the context investigation. There were three main goals for the interview:

- To identify the target user of our research.
- To understand how the problem arises and how users are now completing tasks.
- To understand under what circumstances, they will use the new solution.

The collected results were then used to extract design requirements and expose proper solutions in the next design analysis stage.

3.1.2 Methods

We employed interview techniques according to UCD theory. An interview protocol with the necessary material for discussion was prepared (see Appendix I). According to Jacob and Furgerson [74], this protocol guides the interviewer along the focus of the research, so they can better arrange time for each goal. Literature describing the IR procedure was also used as a reference for the design questions. A picture of the context or related facilities could help both sides communicate detailed descriptions.

We designed an interview protocol with questions according to the literature and interview strategy suggested by Wendy E. Mackay[14]. The strategy for designing questions was, using the proper format to gain the most effective results. Open questions were combined with follow-up questions to better understand the team,

[14] http://insitu.lri.fr/People/AdvancedDesignOfInteractiveSystems2017

context, and working procedure. Setting these questions at the beginning can help the interviewees evoke their knowledge and experience in the practice. Follow-up questions can avoid guiding interviewees in a certain way from the beginning. For example, we asked interviewees the general problems they encountered at the beginning rather than asking them the hands-occupied problem directly. Then we dived into the mentioned issues that relate to hands-occupied problem by a follow-up question. A questionnaire (close-ended questions with an ordered response) gathered subjective opinions on introducing VI into IR.

The interview was divided into three blocks with different topics. The three blocks were designed to adapt to experts with different backgrounds. We asked radiologists questions in all three blocks, but design experts only received questions from Blocks 2 and 3. In the first block, questions about the procedure and teamwork were asked. The second block covered specific interactions, especially those between different roles with different facilities. Block 3 addressed radiologists' subjective opinion of VI. The three blocks also aligned with the goals of the research. The detailed questions can be found in Appendix I.

3.1.3 Participants

Two kinds of related professions were approached, interventional radiologists and design experts for IR facilities. The interventional radiologists provided insight into the real context. Alternatively, the design experts offered insight on technical possibilities and limitations. Our primary purpose was to understand the users, not only what they do but also what they think and feel. A record of the interviews was taken but only with the permission of interviewees. Interviewees had the right to terminate the interview at any time.

Six related experts in total were interviewed, including two interventional radiologists, three professionals in the IR facilities industry, and one with both backgrounds. The first radiologist had 15 years of experience in CVIT in hospital. Another radiologist was an expert in neuro-intervention. They used the C-arm type GE single-arm argo730. As they were remotely located, we used a phone interview with the first two experts, which limits communication when describing detailed interactions.

Three design experts of IR facilities were interviewed by visiting their office. The first was the senior designer of C-arm devices. She introduced the functions of the table console and how it works using a prototype. Another expert explained the design principles of IR and provided significant insight from the perspective of technical feasibility. He was also a C-arm device designer. The third expert was a user experience designer with experience in ASR for the medical system. She suggested the design principle of VI. The last interviewee was an expert who used to be a cardiovascular radiologist. He became a product designer of interventional facilities. We conducted a face-to-face interview with him. He was also the only one who supported an audio record. According to the confidential agreement, we can only publicize information related to our research without mentioning personal information.

3.1.4 Procedure

Before the questioning, a consent form – including the research goal, interview procedure, and confidentiality – was provided (see Appendix I). We then conducted the interview procedure according to protocol.

3.1.5 Results

We collected the results of the following topics from the interview, then integrated those of all the interviewees for better understanding. The original records of each interviewee are in Appendix II.

I. The general procedure for cardiovascular intervention

Cardiovascular intervention is one of the most common intervention types. During the pre-diagnosis, radiologists first identify diseased areas with CT scans or Dual-Source Computed tomography (DSCT) scans. These images can be used to find disease in advance. Sometimes, however, the diagnosis might be wrong. For example, radiologists did not find any disease during the intervention. In a more complex situation, there are three arteries. Therefore, they make the diagnosis by intervention before the treatment.

During an interventional diagnosis, the following procedure is used:

1. First, the radiologist identifies where the disease occurs – the right coronary artery or left coronary artery (LCA).
2. Then, a patient enters the hybrid operating room while the nurse helps with disinfection. The patient then swallows sedatives to relax.
3. The radiologist decides where to insert needles. He/she could use either the groin artery or brain artery when performing a brain intervention.
4. The radiologist cuts a small incision of the vessel and inserts a shift. He/she has the first view of if the shift is correctly placed. Radiologist can see it with the fluoroscopy system.
5. The radiologist next inserts wires to into the left artery. When the catheter arrives (as see under fluoroscopy), he/she begins injecting contrast media. The injection of contrast media in a CVIT is conducted manually. However, in brain or tumor IR, radiologists use a pump. The speed and flow rate can be preset by technologists.
6. After applying the contrast media, radiologists can make a movement of the table, following the contrast media. But this is the seldom case. More often, radiologists already spot if the contrast media is blocked and located where to see.
7. Radiologists sometimes adjust the speed of the injected contrast media. This decision depends on the magnetism of the vessel, its size, and the

required flowing speed. Too much contrast media in the coronary arteries can lead to damage to the heart.

8. Radiologists frequently position the system for optimal placement as the contrast media flows in. They see it from different views: the left artery oblige, the right artery oblige, the spider view, the four-chamber view, and the plane arterial angiography view.
9. Radiologists also adjust the position to find the right pathway. For example, when arriving at an arc but with a view just above it, the radiologist must find an appropriate angle. An assistant can move the table accordingly to provide the radiologist with a better view. Radiologists can see the top of the catheter along with the movement; therefore, the radiologists cannot see where the catheter is inside the body. Radiologists thus use the fluoroscopy to help find the correct pathway in case of a deviation in the artery.
10. Occasionally, radiologists adjust the rate of the frames. Regularly it is 6, 15, or 12 frames per seconds. For different patient conditions, radiologists can increase the speed to 30 or 60 frames for seconds.
11. Radiologists also apply the shunt to create better differentiation inside the coronary arteries and other parts, such as the interference of the lungs.
12. Angiography is triggered by a foot pedal as the contrast media flows in. The created image is shown on a monitor and saved automatically.
13. After all the parts are inspected, the radiologists interpret the captured angiograms.

During the review of the angiograms, radiologists interpret the images to decide where the patient's artery is narrowed. Sometimes the anatomy of a particular patient can be vastly different. For example, the patient may have congenital anomalies. When radiologists consider if it is suitable to perform IR treatment, such as inflating the balloon, placing stents, or increasing vessel dilatation. They should ask for the consent of the patient's family.

During interventional treatment, radiologists perform the angiography again to reach the disease and treat it. They do so to prove that the main stent is optimally placed and fits correctly to the border of the vessel's interior. Radiologists sometimes print the images out to show the patients a comparison of before and after the treatment. Thus, the patient knows they are cured. The radiologists call these related images showing the patients' condition reference images. Sometimes they must inflate the balloon and place a stent inside.

II. The frequently used views

The following screens were needed most frequently: the patient condition monitors (including electrocardiogram (ECG), blood pressure, heart rate, pulse, and blood pressure), real-time fluoroscopy, captured angiography images, reference images (such as the previous CT scan and angiography images), roadmap, and the contrast media monitors occasionally. Notably, a roadmap is different form a reference image. A roadmap is a fusion of the previous image overlaid with the current situation. It is needed to establish a pathway with the best anatomy. A reference image is an old image or CT image created by previous modalities or the old cases.

III. The team composition

The cardio-interventional team usually consists of one radiologist, one assistant, one nurse, and one technologist. An anesthetist or a cardiologist is alternative in CVIT. The anesthetists mostly perform local anesthesia for the patient. In a simple case, it may even be just one interventional radiologist, but it should be a well-trained radiologist in the intervention.

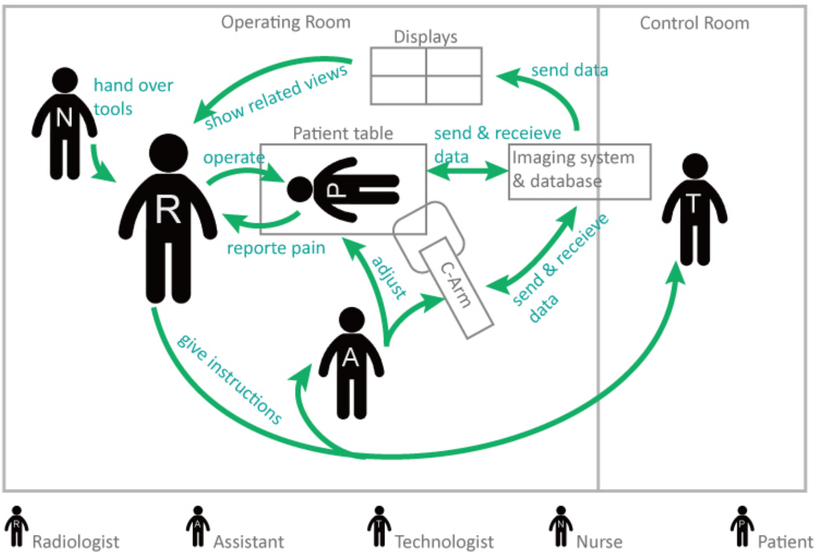

Figure 9: The team cooperation in an IR hybrid room

IV. Cooperation between the main radiologists and other roles

The main radiologists are at the core of the group, as they mainly make decision during the operation. The main radiologist is responsible for inserting catheters and placing stents. The other roles are subordinate and assist the main radiologists with different aspects (see Figure 9). Radiologists cooperate with an assistant most closely when their hands are occupied.

The tasks that radiologists would ask an assistant to help with include the following: adjusting the position of C-arm, intensifier, and table; injecting contrast media or medicine; showing related images; and discuss the patient's condition. Radiologists communicate with them mostly by oral instruction. There are some standard procedures and positions that all radiologic professionals know. Assistants can sometimes anticipate the next requirements without instructions.

The nurses are responsible for handing over catheters; sometimes they also play role of assistant. The technologist sits in the control room. He/she is responsible for setting up and maintaining the C-arm, preparing the contrast pump, inputting patients' name, gender, choosing parameters, clarifying the image, saving images, and displaying reference images for radiologists in the examination room. The main radiologist gives oral instructions to them as well.

Radiologists also ask for cooperation from the awake patient, such as asking them to hold a breath when taking angiograms or move a bit to allow wires and catheters to proceed. Patients should also report their pain during the procedure.

V. The current problem of radiologists

The first problem occurs when giving instructions to an assistant as the radiologist's hands are occupied. The radiologists instead want to manipulate the devices on their own. Though both radiologists and assistants understand the structure of the cardiovascular system, some patients' anatomies are different so the standard position is not suitable. For example, if the patient has a higher weight, their dose is higher and more exploration occurs. The radiologists then communicate with the assistant to re-position the C-arm. In some cases, the position is not optimal. The radiologists then provide more specific instructions like "go a bit more," "come down plus 5, plus 10," "move the table up," or "old position." The radiologist also gives instructions for other tasks – for example, when injecting medicine, they may request "5 milligram medicine," "shunt in," or other instructions such as "reduce the frame." Misunderstanding commands not happen very often. If an assistant does not interpret the requirements correctly, the radiologist would repeat the instructions.

The problem of radiologists moving to the non-sterile control room does exist in practice. Radiologists sometimes enter the control room directly to manipulate devices when the technologist cannot place devices as needed. It also depends on how many radiologists attend the procedure. If there is only one radiologist, this problem occurs more often, at least once per treatment. The sterile requirements are not as strict in the IR as open surgery. There are sterile covers to protect the radiologist, so the problem is not as severe. The radiologists also enter the control room to communicate with patients' family, such as asking for consent to proceed.

Another problem is that radiologists are often very tired after finishing an IR. Largely, they rely on lead aprons and hats that they wear to protect them from radiation. This gear weighs around 10–15 kg and covers the body tightly. Radiologists also sometimes stand to perform IR. The length of the procedure varies from patient to patient, as well as the location of the artery and how complicated the procedure is. The angiography takes around 30 minutes; treatment usually takes 2–3 hours. When they encounter a complicated procedure or challenging cases, the mental load is heavy as well.

The radiologists also mentioned other issues: “The systems are not integrated or function well;” “Nurses sometimes also misunderstand the materials;” and “The high-speed of development of technology is also challenging for them. They need to learn new technologies or be trained to catch up with new trends.”

VI. Table console functions.

A table console is an interface designed for controlling IR systems during the procedure, especially for various positional demands. Multiple modes are designed for better and faster positioning. Figure 10 shows the new control panel in the Siemens Artis phone system. The left sub-panel adjusts the table with a joystick, the signs of manipulation side, and up and down moving buttons, block movement button and emergency stop button. The table can also be adjusted manually forward, backward, left, right, up, and down.

The middle section controls the C-arm and images. The thinner joysticks in the upper-middle are for rotating the C-arm, while the right-most joysticks move the C-arm to a different direction. It can also be locked to moving horizontally or following the direction of the patient table (from head to toes) by switching to the corresponding mode. The image-viewing joysticks are surrounded by several image-adjusting buttons. The block movement, block radiation, balloon time counter, and emergency buttons are placed on the side of the consoles.

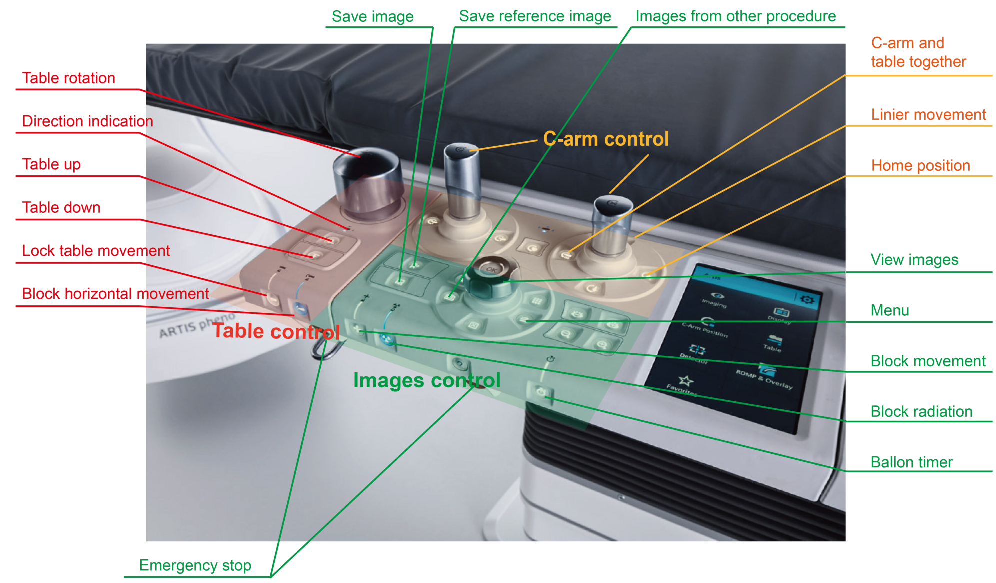

Figure 10: The functions of each button and joysticks imbedded in Siemens Artis pheno

VII. User’s feedbacks to voice interaction

We collected the radiologists’ subjective opinions of independent manipulation by VI at the end of the interviews. We also consulted three user experience experts in the imaging area to understand the technical feasibility and design opportunities of VI. We used a scale with score ranging from 1to 5. The highest score (5) indicates they regard VI as very possible, while 1 represents not possible at all (see Appendix II).

Both the radiologists and design experts expressed common positive feedback in the following tasks: Setup displays (M = 4.67, SD = 0.58); Call an expert remotely(M = 5.00, SD = 0.00); Input contrast parameters (M = 4.25, SD = 0.50); Start/stop contrast injector (M = 4.33, SD = 0.58); Zoom in/out images (M = 4.67, SD = 0.58); Lock/unlock voice interaction (M = 4.67, SD = 0.58); Zoom in/out reference image(M = 4.00, SD = 0.82); View reference image (M = 4.00, SD = 1.41);Adding mask (M = 4.50, SD = 0.71). These actions can seemingly be accomplished with simple and precise commands.

However, the answers from the radiologists (represented by M_R) and design experts (represented by M_D) differed on these tasks: Adjusting the position of C-arm (M_R = 5.00, SD_R = 0.00; M_D = 3.00, SD_D = 1.00); Adjust the table(M_R = 4.50, SD_R = 0.71; M_D = 2.00, SD_D = 0.00); Arrange displays(M_R = 5.00, SD_R = 0.00; M_D = 2.50, SD_D = 0.71). As these tasks could help a radiologist work more independently, but from technical perspective, it is hard to guarantee the accuracy and fluency of such interaction.

All interviewees negatively regarded “capture angiography images”. Both groups described this task more complicated as impossible, with an average grade 1.00 (SD=0.00). This belief reflects the critical timing requirements of triggering angiography with a foot pedal. It is also difficult to describe the position parameter.

3.2 Video observation

3.2.1 Motivation

After the interview, we still lacked several details. For example, how are the hands of the radiologist allocated during medical tasks? How is the working atmosphere? The aim of video observation is thus to discover the detailed interactions missing in the interview.

3.2.2 Methods

Therefore, we focused on the aforesaid ambiguous details. We listed the focus points as follows:

How are the two hands allocated separately?

How exactly do radiologists allocate their two hands, and do they always hold wires and catheters? How focused or alert is the radiologist?

How is the communication?

How do the radiologists, assistants, and technologists communicate during the requested interaction? What are the exact names or parameters of their commands?

3.2.3 Resources

We searched online recourses and targeted video materials containing a real context or real cases in IR. We collected public recourses such as medical case studies, documentaries of interventional radiologists, introductory films by hospitals or medical foundations, and other medical records by individual radiologists who want to share their knowledge.

Two records of a live-demonstration coronary angiogram and renal angiogram through a radial artery by Media Space Plus[15,16] were found. A clear procedure of a diagnosis was demonstrated. The doctor also explained the procedure while working, clearly demonstrating how he allocates his hands and uses the images. The communication among the group is recorded as well. Another live record was Dr. Scott Nowakowski, Dr. Rahul Patel and Dr. Jas Virk performing a minimally invasive uterine artery embolization in a woman with symptomatic uterine fibroids, using a transradial approach[17]. The situation of hands and communication during the procedure provided further details.

Another resource is the introduction video by Lakeridge Health Foundation in Durham Region[18]. It demonstrates several small instruments to treat tumors and introduces a showcase of C-arm movement, including on-focus, over-table, and under-table flexibility. The BSIR[19] uploaded a video that defines IR and explains why IR – or "image-guided surgery" – provides excellent outcomes for patients. It also shows how two hands are allocated when making a puncture, inserting wires and catheters, and injecting medicines. Meanwhile, the interventional documentary by radiologist William Julien[20] explains the principles of treating women's uterine fibroid tumors with IR. Several frames can be used to identify hand cooperation when eliminating the medicine to stop the blood supply. Other materials such as vlogs show interventional radiologists' daily work in hospital, including Dr. Jon and Dr. Chris working at Great Lakes Medical Imaging. Some records of treatment were included[21].

We also collected two videos that not real case records but they show how the C-arm works. The first video demonstrates the flexibility of the C-arm in the Siemens

---

[15] https://www.youtube.com/watch?v=u7V1KeJBHKM
[16] https://www.youtube.com/watch?v=REYkw3-P2KE
[17] https://www.youtube.com/watch?v=1NFDQcQvd2k
[18] https://www.youtube.com/watch?v=w0EQQzIwk-E
[19] https://www.youtube.com/watch?v=1crWP85QRl0
[20] https://www.youtube.com/watch?v=ac_allTZufU
[21] https://www.youtube.com/watch?v=O5IwNioRkh0

Artis Zeego Q in Universitätsklinikum Jena[22]. The second video is a product presentation of a moveable C-arm and floatable design by Siemens in an exhibition[23].

3.2.4 Results

The hands-occupied problem often occurs during the procedure, but only in the diagnosis and treatment procedure, when wires and catheters are used for precise treatment. As shown in Figure 11, a needle is inserted into the artery, then a guiding wire with a catheter is inserted through the needle.

Figure 11: Radiologists are inserting guiding wire via a needle into the patient's vessel[24].

From the live demonstration by Media Space Plus[25], both the radiologist's hands are frequently switching between a sequence of actions, such as placing the needle, cutting a nip, injecting medicine, quickly placing a sheath, injecting medicine again, and inserting the catheter along with the wires into the sheath. Close cooperation with both hands is critical. When one hand is holding the needle, another hand could insert wires or inject medicines. But when the radiologists do not need to make such precise movements, they can place their hands elsewhere, leaving the catheter and wires in a stable place.

We also discovered how radiologists instruct their assistants in the video record of the transradial uterine artery embolization of symptomatic uterine fibroids[17].

> *Radiologist: "so it looks like we are on the right side here now, emmm…."*
>
> *Assistant: "That's looking some resistance?"*
>
> *Radiologist: "Yeah…"*

---

[22] https://www.youtube.com/watch?v=wjM69jQWJDw
[23] https://www.youtube.com/watch?v=OIwAQAB1V_w
[24] https://www.youtube.com/watch?v=1crWP85QRl0
[25] https://www.youtube.com/watch?v=u7V1KeJBHKM

Assistant: "So I get a one-to-one movement so we are just going to see …"

Radiologist: "Ok, ok, we are good."

Assistant: "more wires? "

Not just during positioning tasks, when the radiologists needs related images, they also ask for help with manipulating the images. For example, in viewing the angiography run during the diagnosis stage, the captured image is reviewed immediately after a part of the body is scanned. The radiologist repeats this procedure until they have inspected all possibilities. The position of the captured image is record by the system for further treatment[15]. Figure 12 shows an example of a captured angiogram with exam and sequence order, dose rate, table geometry, and C-arm position.

Figure 12: A sub-screen of captured angiogram with image parameters when viewing images.

The patients are anxious during the procedure. In the live demonstration by Media Space Plus[15], the patient was always awake, occasionally groaning as they felt pain. Thus, another concern of VI in IR is an aware patient. Patients may feel more cautious because of the VI.

3.3 Discussion

At the beginning of the context investigation, we established three goals for this research:

- To identify the target user of our research.
- To understand how the problem arises and how users are now completing tasks.
- To understand in what environment users would employ the new solution.

We can now discuss them individually to assess if we can achieve this goal. We also compare the results of our research with the literature to reach a final conclusion.

3.3.1 The features of the target user

The main radiologist is the target user. Two features of radiologists can be identified from the results. First, radiologists rely on close cooperation with the group. Second, they are very tired, both physically and mentally.

Radiologists cooperated with the whole team closely. This feature is in line with the third features of IR we identified in literature, namely the close interrelated logic among medical actions (see Section 2.4.1). Hübler et al. also drew the same conclusion [10]. The main radiologist is in charge of leading the procedure and managing the entire team. They run the medical treatment by using the related facilities (Section 2.1). However, they cannot handle all the tasks on their own, in many cases; they rely on the team for a successful treatment.

Radiologists cooperate with the assistant most frequently. They may also sometimes instruct the technologists to help with the devices. A nurse often provides the radiologist with any necessary medical instruments. Radiologists also communicate with the patient to hold their breath when taking images. The patient would also report pain during the procedure. Thus, communication is essential and frequent during IR.

The second feature is that users feel tired, physically and mentally. The physical load is largely due to heavy protective clothes (10–15 kg) and long procedures (30m to 2h per case). This finding can also be seen in the literature review [14]. As we thus found, a cardiovascular radiologist treats 2.62 patients on average every day [35]. Sometimes, they stand to perform IR, so it is even more tiring.

Radiologists feel mentally overloaded due to the multiple mental consumptive tasks. First, the injection of contrast media in CVITs is conducted manually. If too much contrast media is injected into the coronaries, the heart is damaged. Second, multiple checks before and during the procedure are conducted to ensure no overlooked diseases. Third, radiologists rely on an indirect view to anticipate the internal structure of the patient. The continuous and real-time trade-off between successful treatment and less radiation exposure or contrast injection happens throughout the procedure. Treatment also requires a careful and delicate operation. Fourth, extra mental consumption is required when encountering unusual patient conditions, such as low patient weight or poor vascular access, thus a more difficult and complex anatomy. Finally, the radiologists keep learning and training. They need to learn new techniques and related technologies to catch up on current trends.

Radiologists are both mentally and physically overloaded; they need a solution to improve their working experience and efficiency.

3.3.2 The hands-occupied situation

Besides the four features we identified from the literature (see Section 2.4.1) the details of hands-occupied situation were collected from the interviews and videos.

The first finding is, radiologists' hands are not always occupied by holding wires, and catheters. But only in the diagnosis and treatment procedure, when wires and catheters are used for precise treatment. Sometimes, they can even leave the wire and catheters aside, but only if they are in a stable place. The intervention procedure can be generally divided into four stages: the pre-diagnose, diagnose by angiography, interpreting the images, and providing treatment [37].

We identified several tasks that radiologists want to do independently, but their hands are occupied. The position task is typically where radiologists encounter this problem. Other tasks such as viewing images, injecting contrast media, adding masks, and using reference images sometimes happen when hands are occupied [11]. The position task is usually combined with contrast media flow [42].

We also explicated the details of adjusting the positions. The position task can be divided into two cases. One follows the moving of wires and catheters to see where they are in the patient's body. After applying contrast media, radiologists can control the imaging system following the contrast media. This case is less frequent. Most often, radiologists already know where the contrast media is blocked and where they need to see, but they require multiple angles and a slight adjustment to capture the optimal images. Thus, radiologists can adjust the flexible C-arm, the multi-tile table, or the intensifier.

Radiologists position the C-arm from many different views [9]: left artery oblige, right artery oblige, spider view, four-chamber view, and plane arterial angiography view. These present the optimal interpretation for how the contrast media flows. The floating table permits the positioning of the patient, and it can be adjusted manually forward, backward, left, right, up, and down to allow fast and precise movements during IR. The telescoping pedestal is ideal for a surgical operating room with unobstructed neck-to-toe imaging and rapid horizontal panning [8], [9]. Sometimes radiologists adjust the rate of the frames, such as increasing the frame rate of the acquisition system from the more common 6, 15, and 12 frames per second to 30 or 60 frames per seconds (if the system is capable).

Furthermore, when radiologists need related images, they ask the assistant to help as well. For example, in viewing an angiography run during the diagnosis stage, the captured image is reviewed immediately after a part of the body is scanned. The position of the captured image is recorded for further treatment[15]. They repeat this procedure until the radiologist has inspected all the views. Radiologists can either use the table console beside the table or ask the technologist to play the captured images. But they preferred viewing the images by clicking through each frame [11]. Another task during a hands-occupied situation is editing the reference images while administering treatment. These tasks can be completed by instructing the assistant, but the intentions of a radiologist are hard to express precisely in a short time. Even the manually adjustment of DSA helps radiologists interpret images more clearly and effectively (see Section 2.1.2).

The current method of dealing with this problem is to assign an assistant. However, communication is not always fluent and easy. When radiologists produce a wanted position, they tell assistants. There are some standard positions that all radiologic talents know, so sometimes assistants can adjust to the position without communication with radiologists. When encountering an unusual patient condition, continual communication between the radiologist and assistant happens.

In general, we noticed that the requirements of independent manipulation result from different, deeper needs. For example, radiologists want to position the fluoroscopy independently for quicker acquisition and less distraction. They know what they want, but they do not want to continuously repeat their requests. Radiologists manually adjust the mask for DSA image for a better understanding of the results, which helps in the interpretation of artery pathology. Radiologists also want to view the images on their own, as it helps interpret blood flow. Related solution that already allows independent interaction is capturing images with a foot pedal. There is a critical need for more precise timing. Results of the users' subjective opinions on VI in specific sub-tasks can illustrate the differences in independence requirements.

3.3.3 The environment

The IR context is complicated, as multiple roles, techniques, devices, and sub-tasks are integrated (Section 2.1). From the video records, we observed a multimodality interactive system working together to allow the user precise control. Input devices include the joysticks and touchscreen in the table console, the foot pedal, the mouse and keyboard, the computer console in the control room, and buttons in related devices. Still we argue that, a considerable fusion of VI solution and existing interaction modalities in IR context is vital.

3.3.4 Limitation

A significant amount of medical terminology was used during the interview, especially when mentioning a medical apparatus or disease type. Mis-hearing happened occasionally. Moreover, when asking for the users' subjective feedback on VI, we offered several protentional voice commands to help the interviewee imagine the interactions. But we found this text may have led the interviewees and limited their imagination of possibilities. We also found that personal experience influenced the results – for example, designers regarded the tasks unfeasible when the interactions appeared complicated or featured continuous input. Radiologists, however, seemed positive when such an idea occurred to them. But radiologists gave negative feedbacks with the consideration of practical issues like timing, which is more valuable for our research.

As we did not have the opportunity to observe a real IR operation, though we encouraged the participants to describe a detailed story, we still lacked the atmosphere, scene, and some specific interactions.

3.4 Conclusion

In this chapter, we conducted interviews and video observation to understand the users and context. Results regarding the general procedure of IR, the consoles and needed views, the team composition and cooperation, the current problems, the details of positioning tasks, and the user's feedback of VI were collected.

Two features of the target user, namely the main radiologist, can be identified from the results. First, radiologists rely on close cooperation with the group. Second, they are very tired, both physically and mentally. Radiologists' hands are not always occupied by medical tools, only when precise treatment is needed. We identified several tasks that radiologists want to do independently, but their hands are not available. Adjusting positions is one of the typical tasks. We also explicated the details of how radiologists requiring different positions. Furthermore, when radiologists need related images, they ask an assistant to help as well. We also noticed that the requirements of independent manipulation result from different medical needs. Each task needs to be analyzed respectively. The IR context is complicated, as multiple roles, techniques, devices, and sub-tasks are integrated. A considerable fusion of VI solution and existing interaction modalities in IR context is vital.

Related works inspired us such that VI might be a possible way to assist our users' autonomy. However, both of the IR context and the procedure are complicated. The user's state and team cooperation need to be taken into account when introducing VI. We collected details of the hands-occupied problem. However, they are too descriptive and complex. It is still challenging for us to identify design opportunities and challenges. Therefore we adopted a synthesize stage according to J. Zimmerman et al. [25] to extract design requirements and expose proper solutions in the next chapter.

4 DESIGN ANALYSIS

In the previous chapter, we presented the results of the interview. Topics such as general procedure, team cooperation, target users' state, and working context were introduced. We also compared our results with existing literature to foster a more comprehensive understanding. However, the collected data are still complex and detailed. We referred to a design process proposed by J. Zimmerman et al. [25] to help synthesize the gathered information. In this chapter, we first disintegrate the complicated hands-occupied problem into small pieces (see Section 4.1) to extract design requirements (see Section 4.2). We also employed the UCD design methods such as personas and scenarios (see Section 4.3) to identify design opportunities and challenges (see Section 4.4). Our ultimate purpose is to find a solution that can solve the hands-occupied problem, generating further benefits from better performances and user experiences.

4.1 Defining the problem

Both the hands of the main radiologist are often occupied in several tasks. However, when is this situation a problem? "Occupied hand" is only a situation; it only becomes a problem when four conditions are met. We propose these four conditions to better understand the problem and further specify user requirements and potential solutions.

The first condition is when radiologists encounter limited flexibility, especially gesture flexibility, because of one task. We call this task the “primary task.” We specify this condition, because the two hands are not fully occupied throughout the entire procedure. When no delicate manipulation of the primary task is required, radiologists can free one hand to do other tasks. Sometimes, radiologists can even leave wires and catheters set aside, only ensuring the wire and catheter are in a stable place. Delicate and precise manipulation exists mostly during the medical stage, when the radiologist is attempting to find the best pathway, diagnosis a disease, or run a target treatment. The radiologist cannot be replaced for this primary task; as such, their hands are occupied.

Second, a “secondary task” occurs when the radiologist’s hands are fully occupied. This task is namely the targeted task for introducing VI. We specify this condition because some secondary tasks may be accomplished before or after the primary task. For example, when editing reference images, radiologists may do so before or after the procedure. It would be ideal if radiologists could make real-time editing when doing a precise placement. Then it brings the third condition, is it necessary to empower radiologists to do all the potential existing secondary tasks?

The third condition is the radiologist's desire to accomplish the secondary task on their own. This condition also aligns with UCD theory: design results should match the users' needs, not only the designer's creativity. From the interviews, we noticed that the requirements of independent manipulation were caused by different deeper needs. Therefore, we argue only the tasks that the radiologists regarded as a problem are the focus of our research.

The fourth condition is that radiologists cannot address the secondary task with their own hands in the current situation. A typical example that does not fit the condition is, radiologists must capture angiograms while ensuring the precise placement of wires and catheters to inject contrast media. A foot pedal is offered to solve this problem. We address the word "cannot", as it could result from limited hands flexibility or limited cognitive resource. The specific reasons that prevent radiologist's autonomy is critical.

4.2 User analysis

4.2.1 User's mental model

Before we define the users' requirements, it is important to understand their mental model when dealing with two tasks.

The first issue is the priority of the two tasks. The primary task has highest priority when radiologists are facing multiple tasks, and it also decides the capacity remaining for the other task. Radiologists spend both their physical and mental resources on the primary task, as it is the key task for a successful treatment. In the current situation, though radiologists are not solely responsible for completing the secondary task, they must instruct another person, which also requires cognitive resources. From the interviews, radiologists are responsible for the medical success and management of the group. All other team members attend to help the procedure. Ensuring medical success is, thus, their first goal.

How the user's capacity is occupied by the primary task determines the capacity remaining for handling the secondary task. The Baddeley & Hitch [61] working memory theory noted that working memory has a limited capacity. Lavie's [75], [76] cognitive theory also proposed that the cognitive resources used for primary tasks determine how they use the remaining cognitive resources to process the other stimulation. The complexity and delicateness of the primary task influences the performance of the secondary task. Thus, we find that the severity of the two-hand-occupied problem varies in different cases. In some simple situations, radiologists can manage the entire operation independently, even injecting the contrast media. But this independence is only in simple diagnosis cases. For most cases, the radiologist needs help from assistants.

This resource distribution model helps us understand why radiologists cannot take care of two tasks at the same time. First, their hands are occupied by the primary task, so radiologists distribute their physical resources to that task. Second and less obviously, they do not have enough cognitive capacity left for the secondary task; radiologists have a limited capacity while working. The new solution should not only tackle the problem of occupied hands but also fit within cognitive capacity or help reduce cognitive load.

To summarize, the mental model of radiologists can be described – according to the information provided by the different facilities, other roles, and their own feelings on hands, a radiologist decides what to do next and then arranges resources accordingly. Therefore, we regarded the success of the primary task as the dominant requirement of the user, while the independence and success of the secondary task are subordinate.

4.2.2 User requirements

As radiologists cannot be replaced in their primary task, a delicate and precise manipulation is necessary. The secondary task wherein radiologists require independent manipulation is the target opportunity for VI. We aim to empower radiologists with the independence to run secondary tasks by VI rather than relying on an assistant.

We defined the following user requirements:

- Solutions shall enable radiologists complete the first and secondary task at the same time, independently.
- Solutions shall enable radiologists complete the secondary task without hindering the primary task (not resulting errors in the primary task).
- Solutions shall enable radiologists achieve better performance (fewer errors and faster procedures) with the solution.
- Solutions shall enable radiologists have better experiences (better satisfaction and less cognitive load) with the solution.

However, the hands-occupied problem is complicated, as different tasks have their own interactions and medical needs. The reasons that currently limit their autonomy, are the challenge of our research as well. When exploring the solutions, the radiologists' allocation of their physical and mental capacity to specific tasks must be clarified. A more concrete and contextual interaction narrative is needed. Such UCD methods as persona and scenario were used to find where and how VI could be introduced.

4.3 Persona and scenario

Persona is a typical UCD method that synthesizes the characters of the target group from interviews to create a character. We also introduced a series of secondary personas – an assistant, a technologist, a nurse and a patient – in our research, as they all cooperate to ensure success in the procedure. A scenario is a story demonstrating how the main character cooperates with the related roles by conducting a sequence of events to achieve their goal. Both the persona and scenario are built according to the users' inputs in the interview. There are three types of scenarios: best-case, worst-case, and average-case. These types refer to the conditions described in the scenario as optimal, worst, or average for the main characters in achieving their goal. As an exploratory study, we used the average-case scenario to avoid extreme results.

4.3.1 Persona

The key persona is the main radiologist, who is also the main character in the scenario. **Radiologist Kevin**, 45 years old, has 20 years of experience in CVIT. He is American and works in a public hospital in Amsterdam. He is basically healthy but suffers from a chronic lumbar muscle strain from years as a radiologist.

There are three secondary personas. **Assistant Lily**, 25 years old, has been working as a junior interventional radiologist for 2 years. It is her fourth time as an assistant for Dr. Kevin. **Nurse Amy**, 30 years old, has 5 years of experience as an interventional nurse. **Technologist Peter**, 35 years old, previously worked as a CT technologist for 8 years. But 3 years ago, he decided to divert to interventional technology. **Patient Mary** is 46 years old. It's her first time undergoing a cardio-intervention. She has a family history of heart disease, but this is the first time she has found problems in her coronary artery.

4.3.2 Scenario

Today, Kevin has already completed 2 operations; he is slightly tired now. He knows patient Mary, as he is her attending, and he completed her CT pre-diagnosis with her last week. Two potential stenosis in the LCA were marked, and a reference image indicating the position of the stenosis was stored in the database. Kevin arranged the angiography intervention last week to recheck the diagnosis before interventional therapy. Consent from the patient was already achieved. He requested one assistant, one nurse, and one technologist to help with the operation; technologist Peter, Assistant Lily, and Nurse Amy agreed to join. He knows them very well from past cooperation.

Before the diagnosis procedure begins, all the team members (except the technologist) wear sterile aprons, masks, gloves, hats, etc. Before starting the procedure, all the required instruments and medicine facilities are prepared. First, Peter had helped set up the facilities, then sat inside the control room. The real-time fluoroscopy and marked CT images are displayed before the team members. The ECG and blood pressure are monitored, too. A set of protective polyethylene shields cover the equipment.

Kevin has already inserted the wire and catheter into the groin artery. Lily is standing beside the table console to show the corresponding fluoroscopy. The wire and catheter are in the center of the view. Amy is standing by. She delivers Kevin the required instruments. Mary is awake. Though she was thoroughly told about the procedure, she is still a bit nervous. She was told that she should report any pain to Kevin. As Mary has a family history of heart disease, injecting the contrast media and inserting wires should be done carefully.

Kevin inserts the guide wires into to the narrowed vessel in the LCA. Though Kevin can only see the visible wire in the real-time fluoroscopy (the vessel is invisible without contrast media), he moves the wires and catheters up based on his knowledge and experience with human anatomy. Meanwhile, Lily moves the position of C-arm accordingly to ensure the wire is clear in the view. Kevin moves wires fast, but as he is very tired, he accidentally hits the vessel a bit, and Mary groins immediately. Kevin comforts Mary and says: “Painful? Sorry, don’t worry, it is fine.” Mary nods in response.

When Kevin arrives at the entrance of the LCA, he decides to inject some contrast media to highlight the narrowed vessel. He asks Lily to start the injection pump for the contrast media, and Mary holds her breath as he takes angiograms with the foot pedal. Lily injects the standard dose of contrast media at a standard speed, and the contrast media flows through the LCA. In the meantime, Kevin says, “hold your breath!” He also steps on the foot pedal to start capturing, stopping when the full flow is captured. Kevin also tells Mary, “Ok, it’s done.” The first sequence of angiograms, named Exam 1, is stored automatically.

Kevin does not find a narrowed vessel during the real-time angiograms, but he wants to confirm first. He immediately asks Peter to display the first angiograms for reviewing. “Peter, may I review the image now? Thanks.” Peter then arranges a sub-screen to show the captured Exam 1. Truly, there is no stenosis in Exam 1. But it could be hidden because the angle is poor. Usually, they take multiple angles. Kevin and Lily both know they must take angiograms from other angles; therefore, Lily adjusts the C-arm to another standard position without instructions. They repeat the procedure and take 3 more positions like the first one.

Now, Kevin wants to review all the captured images. He does not need to hold the catheter and wire; just ensure they are at a stable place. This time he checks each exam individually by using the joysticks on the table console. One blurring stenosis in a tiny vessel is found in Exam 3, but it is not clear enough. Thus, Kevin tells Lily, “We need to go further.” Lily nods.

To find the right branches, Kevin asks Peter to display Exam 3 as a reference image. Peter then clicks through the frames in Exam 3 and asks, “Which frame do you want?” Kevin views the sequence and when he finds the proper one, he immediately says, “Stop, this one please.” But Peter has already clicked several more frames over, so Kevin instructs him, “emm... back, back, back... yes, this one, please, thanks, Peter!” Then Kevin inserts the wires and catheter further inside the artery according to the reference image, but he finds that the current angle of the C-arm is not optimal, so he says, “Lily, can you please rotate the C-arm to the left, around 15 degrees?” Lily replies while adjusting, “like this?” Kevin then provides feedback according to the real-time position, “more, more, more …ok, stop here, yes, this angle is good” before he inserts wires further.

4.3.3 Interaction flow

We noticed that multiple interactions happened during the scenario.

Table 1 illustrates the interaction flow by specifying the roles and modalities. We thus analyzed how the characters interacted in the primary and secondary tasks, respectively.

Table 1:The interaction flow and modality in scenario.

Actions	**Radiologists (Target user)**		**Assistants/ Technologists**	**Input device**	**System Feedback**	**Interaction Modality**
	Modes used in primary task	Modes used in secondary task				
Insert wires to LCA	Hands for inserting wires Eyes to see the wires in fluoroscopy		Moving the position of C-arm accordingly	Joysticks	C-arm movs Corresponded image shows	Tactile Visual
Inject contrast media	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Voice for telling assistant the inject contrast media	Hearing the instruction Asking questions when needed hands for injecting contrast media	Contrast injector	Corresponded screen shows	Tactile Visual Auditory
Capture angiograms	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Foot to step on foot pedal Voice to instruct patients		Foot pedal	Recording the image when foot pedal is on Automatically saving captured image	Tactile Visual Auditory
Review captured images	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Voice for telling technologist display the captured image	Hearing the instruction Asking questions when needed Arranging screens Displaying images	Mouse and keyboard	Captured image shows on another sub-screen	Tactile Visual Auditory
Move to another standard position	Hands for keeping the wire stable Eyes to see the captured images		Moving to the standard position without instructions	Joysticks	Corresponded image shows	Tactile Visual
Capture angiograms	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Foot to step on foot pedal Voice to instruct patients		Foot pedal	Recording the image when foot pedal is on Automatically save captured image	Tactile Visual Auditory
Loop “move to another standard angles” and “capture angiograms” two more times						
Review all captured images (4 in total)	Keeping the wire stable. Hands are not necessary.	Hands are used to view the images. Eyes to see the captured images		Joysticks	Displaying exams one by one	Tactile Visual Auditory

Actions	Radiologists (Target user)		Assistants/ Technologists	Input media	System Feedback	Interaction Modality
	Modes used in primary task	Modes used in secondary task				
Insert wires further	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Voice for telling the group next step	Hearing the instruction Asking questions when needed Moving the position of C-arm accordingly	Joysticks	C-arm moving Display corresponded image	Tactile Visual Auditory
Ask for a reference image	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Voice for telling technologist to display Eyes for choosing the right image	Hearing the instruction Asking which image Clicking through frames in Exam3 Selecting and displaying reference image	Mouse and keyboard	Show each frame in exam3 Displayed the selected image are	Tactile Visual Auditory
Adjust the angle of C-arm	Hands for keeping the wire stable Eyes to see the wire in fluoroscopy	Voice for telling assistant the position they want.	Hearing the instruction Asking questions Moving the C-arm according to instructions	Joysticks	C-arm movs Corresponded image shows	Tactile Visual Auditory

4.4 Solution discussion

4.4.1 Opportunities

According to the problem definition summarized in Section 4.1, we first analyze the four conditions of each task (see Table 2). A discussion of the design opportunities and challenges follows.

During most of the interactions, both hands of the radiologist were occupied with holding wires and catheters. Only when the radiologist finished capturing images and began interpreting them did the radiologist free his hands. A secondary task exists simultaneously with the primary task. We aim to empower radiologists with more autonomy so they can perform these secondary tasks themselves without relying on an assistant. These secondary tasks are the target task for applying VI. The design opportunities employ VI to complete the secondary task by replacing the original interaction of providing instructions, inputting data to the system via an assistant until the system offers feedback. Some designs already allow the independency of the radiologist in addressing both tasks – for example, using a foot pedal to capture angiograms. The desire for independence was compared with the subjective feedback from VI during the interview (see Section 3.1.5).

Table 2: Analyzing each task with the four problem conditions.

Actions	There is a primary task	There is a simultaneous secondary task	Desire of independency	Reasons of limitation	Conclusion
Insert wires to LCA	✓	✓	✓	Hands are occupied by the primary task.	✓ Regarded as opportunity
Inject contrast media	✓	✓	✓	Hands are occupied by the primary task. Focus on interpretation of anatomy.	✓ Regarded as opportunity
Capture angiograms	✓	✓	Independency achieved by foot pedal	Hands are occupied by the primary task. Focus on timing the capture.	
Review captured images	✓	✓	✓	Hands are still occupied by the primary task. Focus on interpreting images.	✓ Regarded as opportunity
Move to another standard position	✓	✓	✓	Hands are occupied by the primary task.	✓ Regarded as opportunity
Capture angiograms again	✓	✓	Independency achieved by foot pedal	Hands are occupied by the primary task. Focus on timing the capture.	
Review all captured image (4 in total)	Hands are not necessary to be occupied	✓	✓		
Insert wires further	✓	✓	✓	Hands are occupied by the primary task.	✓ Regarded as opportunity
Ask for a reference image	✓	✓	✓	Hands are occupied by the primary task. Focus on choosing a reference image.	✓ Regarded as opportunity
Adjust the angle of C-arm	✓	✓	✓	Hands are occupied by the primary task. Focus on identify optimal angles.	✓ Regarded as opportunity

Thus, we regard the following tasks as design opportunities for introducing VI:

- Moving the C-arm and the inserted wires and catheters
- Setting and adjusting the parameters of the contrast media with an injection pump
- Asking for the display of captured images during the diagnosis
- Moving the C-arm to a standard position when diagnosing disease
- Slightly adjusting the angle of C-arm to recognize the path of arteries.

4.4.2 Challenges

From this scenario, we identified two challenges when applying VI in IR. The first challenge is the distribution of cognitive resources. In other words, do users still have capacity to handle VI? Reasons the radiologists may be blocked from performing a secondary task are specifically discussed in the context of this scenario (see Table 4). The first and main reason is that their hands are occupied by the primary task.

We also found cognitive load during several tasks. Radiologists are focused and dynamically distribute their cognitive resource to multiple agents. For example, when taking angiograms, their hands are precisely placing the wires and catheter. At the same time, they see the contrast media entering and decide the best time to trigger the capture. According to the mental model of radiologists in Section 4.2.1, the cognitive resource used for the primary task determines the remainder for the secondary one. Cognitive capacity is dynamic and hard to quantify; therefore, it is difficult to say if the cognitive capacity left for the secondary task is sufficient.

Whether VI can improve cognitive load or make it worse depends on the specific situation. Prior studies demonstrate that voice-centric systems can sometimes improve performance by using both perception channels. However, it can also result in cognitive overload when too many channels are provided or tasks are too complex. (see Section 2.2.2) In Baddeley's model [61], the central executive can process information from the phonological loop and visuospatial loop in parallel, showing better results in learning when both of the loops are used rather than only one.

Their knowledge and experience of medical treatment are stored in long-term memories and are used to guide the whole procedure. Episodic buffer records information stored in short-term episodic memories to help make decisions. Can VI reduce mental consumption compared to an assistant in such a complicated situation? Though an independent system handled by radiologists is ideal for our users to fulfill their requirements of independent manipulation, VI cannot fully replace the assistant at this time. For complicated interactions – such as drawing lines on a reference image – it is hard to only rely on VI. We also found that the assistant was not replaceable in some complex procedures, which means they still spend their attention on multiple targets.

The second challenge is background noise, especially when someone asks the user questions. This problem is related to the gratuity of the voice system as well as the safety and accuracy of VI. A VI solution should not cause chaos in team cooperation and group communication. The other roles must speak to meet the requirements of a successful treatment. However, the system can be designed to recognize a target voice, thus avoiding the chaos of multiple speakers. A speaker-dependent system is preferred.

However, conflicts still exist when a radiologist wants to use a voice command while others are speaking. Though people usually talk only when necessary, the team members should avoid talking when radiologists giving out commands. One issue not demonstrated in this scenario is the patient randomly groaning or reporting pain during the procedure. How would the voice command overcome overlap with other noises? Or when the input procedure is interrupted by accidents? The system could use advanced technologies to eliminate background noise, and it must provide the user with clear feedback on its current processing stage. Users must know where they are and what to do next. Furthermore, to increase accuracy, the database should be restricted to a limited vocabulary and the users trained in advance.

As the cognitive issue is critical for meeting safety and efficiency requirements, VI is the best option to apply in interactions when the voice command can clearly and precisely deliver input data. A sub-task – like adjusting the C-arm to the standard position – is possible. Reviewing captured images from the database is possible as well. Continuous interaction – such as moving the C-arm or reviewing a sequence of images – deviates from the efficiency requirement but follows the independence requirement.

To summarize, two challenging conditions were identified. The first challenge is the distribution of cognitive resources. When the primary task is complicated, for example, an unusual patient anatomy appeared or when radiologists are asked questions. The second challenge is background noise, especially when someone asks the main radiologists questions. A study is imperative for finding out the answer out research question.

4.5 Conclusion

In this chapter, we discussed the design and introduction of VI in IR. We first divided the hands-occupied problem into smaller factors and discussed our conditions behind the problem. From this discussion, the mental model of a radiologist dealing with two tasks was analyzed. We also explained the user requirements for further solution exploration. A set of personas and a scenario were then used to construct a more concrete and contextual interactive story, and we analyzed the interactions of this scenario. Five design opportunities and two challenges were discussed, respectively.

However, whether VI is a solution or results in greater chaos relates to the context and primary task. Can VI solve the hand-occupied problem and also improve working performance and user experiences? We must test to reveal these results. In the next chapter, we introduce a study to collect the results.

5 STUDY

In the previous chapter, we discussed the opportunities and challenges of introducing VI into IR. The possibility was uncovered as radiologists' autonomy can be reinforced by a voice control system. However, when the context or tasks are complicated, VI may distract radiologists inversely. In this chapter, we designed a simulated study to ascertain the effect of introducing VI into IR, especially when a challenging condition appears. We used methods to measure working performances and user experience [26]. We present the results and further discussion at the end of this chapter.

5.1 Motivation

The IR context is complicated as there are multiple roles, techniques, devices, and sub-tasks integrated together. We identified five possible design opportunities (see Section 4.4.1); however, their effect on performance improvement and user experience is still doubtful. In challenging conditions, introducing VI may result in more errors.

The primary aim of the study is to determine whether VI improves radiologists' performance and experience in challenging conditions. During the study, these variables were measured and compared under challenging and non-challenging conditions. Two challenging conditions were identified in Section 4.4.2: when the primary task is complicated, for example, an unusual patient anatomy appeared or when radiologists are asked questions. We ran the study in a simulated context due to resource limitation. We could not run the study in a real operating room with radiologists; therefore, we opted for a study, focusing on human beings who work as radiologists.

To summarize, the motivation of the study is to ascertain: how challenging conditions influence work performance and user experience in a simulated context where VI is offered**.**

5.2 Context simulation

The simulation involves radiologists performing the task of inserting wires and catheters into the arteries. Because the IR procedure is complicated and entails multiple facilities. Simulating the facilities and the environment of a focused task is more feasible. The simulation was divided into three aspects: simulating the way of conducting IR, simulating the devices and technologies used to support the procedure, and simulating the working environment.

5.2.1 Simulating the way of conducting IR

The way of conducting IR was simulated by playing a game "Spiral Action Game" (see Figure 13). The original rule is to move the hook from the start to the end without touching the metal route. It is a pathway game that involves precise placement. This procedure is similar to the way radiologists insert wires and catheters through an artery. Both procedures require a combination of tactile feeling and visual information to ensure precise placement.

Nonetheless, conducting IR and playing a spiral game are different in two ways: First, only one hand is needed to play the game, while radiologists use both hands to perform IR. Second, the angles needed to interpret the wires in the arteries and the hook around the metal path are different. We proposed two solutions to address these differences. First, we asked participants to use their dominant hand to perform the task and keep the non-dominant hand at a standstill. Second, although the specific angles of view needed are different, we focused on simulating the moment when radiologists need other perspectives to ensure precise position rather than simulating the exact same angles of view.

Figure 13: A child is playing the spiral game

5.2.2 Simulating related devices and technologies

Related devices used during the insertion of wires and catheters are mainly real-time fluoroscopy and displays. Radiologists combine reference images and real-time fluoroscopy to interpret anatomies. When arterial structures are hard to recognize, another angle of fluoroscopy is displayed, or captured high-resolution angiograms are used as references (see Section 3.1.5).

Figure 14: A participant playing the spiral game in the simulated context and setup

To simulate this procedure, the spiral game is placed inside a box (see Figure 14). Participants cannot see the game device directly. Four views—a real-time front view (F), a real-time left view (L), a real-time top view (T) and a static front view (M)— are displayed in two screens to guide the game. The static front view is the path map (see Figure 15). Three cameras were used to present the real-time situation. Figure 16 illustrates the testing setup.

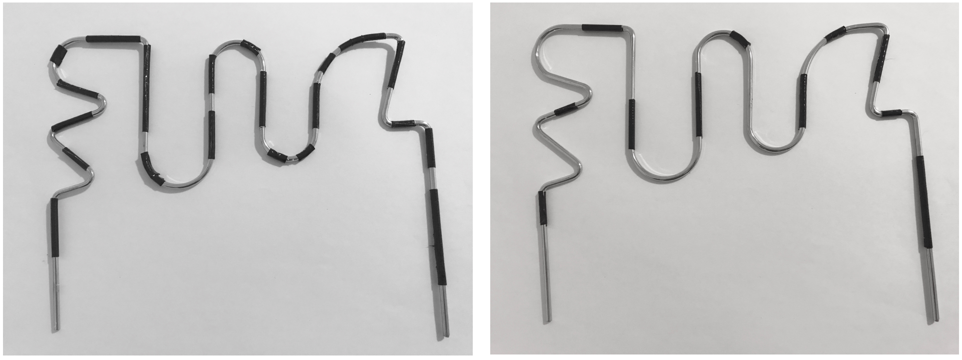

Figure 15: The static front view shown to participants. The left map with more isolated tapes is the standard route used in tests B and C. The right map is the challenging route used in test D.

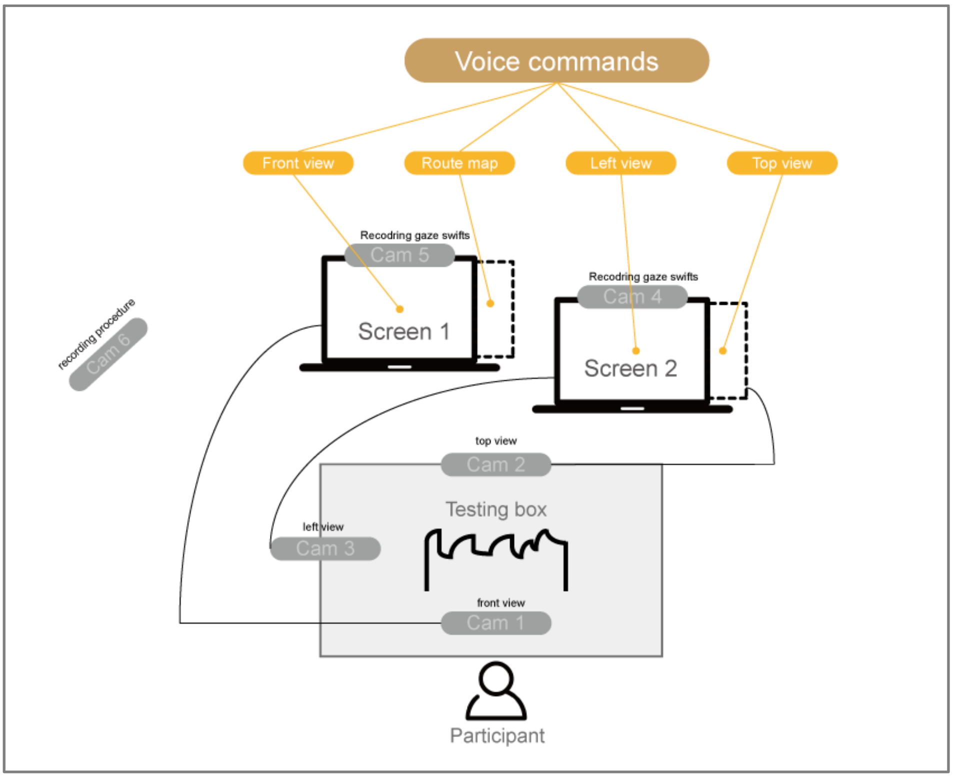

Figure 16:The testing set-up used in the study.

By default, the top view and the map are always displayed in two separate screens to simulate real-time fluoroscopy and a reference image. The top view can be switched to the left view when needed. Participants can also switch the view back. This procedure simulates the need for other angles during IR. Different perspectives are also required to play the game when the hook goes through a horizontal or a vertical path.

Usually, the top view and the map were always displayed in two sperate screens to simulate the real-time fluoroscopy and a reference image. The top view could be switched to left view when needed. Participants could also switch the view back. This procedure was simulating the need of another angles during IR. Different perspectives were also required to play the game when the hook went through a horizontal or vertical path.

The real-time front view was designed to be displayed only for a limited duration (5 seconds) because when radiologists take high-resolution images, they also limit the usage to avoid radiation exposure. During the game, a real-time front view is useful as participants can clearly see the position like high-resolution images in IR. The fixed time duration was pre-defined by our prior testing in case participants remain in the real-time front view and not switch back. Though we did not find reference information about how long an image capture usually last, 5 seconds are just sufficient to identify the condition clearly.

5.2.3 Simulating the environment

In addition to simulating tasks and devices, we also simulated the environment based on background research (see Section 2.4) and user interviews (see Section 3.3.3). We simulated user conditions by requiring participants to wear their winter jacket to simulate lead aprons. Although winter jackets weigh less than real lead aprons, it is helpful to mimic the feeling of restricted flexibility. Moreover, the fixed time duration of playing the game resembled the time pressure radiologists experience in an understaffed situation. Table 3 outlines the simulation in detail.

Table 3: Aspects of the simulation study.

Items	In IR	in Test
Tasks	Radiologists insert the catheters and wires through the arteries precisely and delicately.	Participants carefully hold the hand hook through the metal route.
Items	In IR	in Test
Pathway	A wire through a three-dimensional artery with pinpoints of disease	A two-dimensional hook though a two-dimensional metal route with safe areas covered by isolation tapes
Error feedback	Radiologists realize errors by tactile feeling and visual feedback. They feel wires and watch real-time images together to see if wires touch the wall of vessel.	Participants realize imprecise placement when the warning alarm goes off each time the hand shank touches the metal route.
Environment	Hybrid operating room (multi-role, multi-devices, bright)	Open and bright room in the DesignLab. The environment is open to other people. Only the host can talk to the participants according to the testing plan.
Necessary view	Two image screens: One for real-time fluoroscopy to see the wire and	Two screens: One screen displays the front view as a routemap, while another screen

	another for a reference image (old captured high-resolution image when contrast injected and later processed - DSA) to show the path of artery	displays the left view or the top view to allow participants to see the hook (voice controls which view is displayed for precise position)
Physical state	Radiologists wear a lead apron.	Participants were asked to wear a jacket.
Time pressure	As fast as possible without compromising safety	Participants were asked to go as far as possible, triggering less alarms (prior), in a fixed time duration.
Both hands occupied	In IR, radiologists usually use the dominant hand to insert catheters or inject medicines and the non-dominant hand to hold the entrance needle.	Participants use the dominant hand to hold the hand hook while the non-dominant hand remains at a standstill.
Displays	A large display with four to six sub-screens	Two PC monitors

5.2.4 Defining voice interactive rules

A mimic voice control system was used to ascertain the effect of VI in different conditions. Participants can request the four offered views with clear voice commands. Participants should say “top view,” “left view,” “right view,” or “map/route map/static view” to see the corresponding display. To optimize user experience, commands with clear indication, such as “the last view,” “I need the top,” and “top view and the map” can be successfully processed. Ambiguous commands, however, such as “give me that view” or defected commands, such as “show other views” cannot be successfully interpreted. The system then provides intelligent feedback, such as “please say the name of the view” to guide participants.

A wizard mimicked the VI system and remotely arranged the four views. The wizard, seated behind the set-up, used a wireless mouse or a touchpad to control the views. On the screen, windows were displayed as full screen. When the user speaks out a voice command, the assistant minimizes or maximizes the corresponding window. Participants were required to practice with the system before the tests for familiarization.

5.3 Study design

5.3.1 Control variable test design

The control variable method was used to conduct a within-subject design. The study was divided into four tests. The first was a prior test A that recorded performance when direct views were offered. After which, we compared the three conditions—test B in the standard condition, test C when interrupting questions were asked, and test D in the challenging route condition—with the simulated setup. The standard condition was the condition without the two challenging issues.

Test A: a prior test (Direct view/ without voice commands)

Participants played the spiral game as normal, which means they do not need the simulated setup. They could see the metal path directly without cameras. They also did not use voice commands.

Test B: standard condition (Indirect view / with voice commands)
Participants played the game with the testing setup that only offered an indirect view. Test B is the standard test that would be compared with other tests ran in challenging conditions. Participants used voice commands to call for corresponded views to make a successful placement.

Test C: interrupting questions condition (Indirect view / with voice commands/ with interrupting questions)
Participants played the game with the testing setup like the test B, but were asked interrupting questions while doing so. Participants should try to keeping moving while answering these questions. We asked three questions randomly. The first two were questions regarding the participant's background, such as "Where is your hometown?" or "When is your birthday?" The last question required calculation, such as "What is the answer to 19+28?"

Test D: challenging route condition (Indirect view / with voice commands/ with challenging route)
Participants played the game with the testing setup as the test B. Additionally, a more challenging route with less isolated tapes was used to simulate the complicated anatomy of patients (see Figure 15).

We compared the results from tests C and D with test B to ascertain the effect of the challenging condition. Test B was compared with test A to determine how performance changed when participants cannot see the game device directly.

5.3.2 Game rules

The rules when playing the game are the same in the four tests. Participants move the hook from the start to the end without touching the metal route. The starting point is the end of the metal path on the right side, but for participants who are left-handed, they start from the left side. If participants arrive at the destination before the time is up, they can turn back; the new starting point then switches to the other side. The time duration for each test is 2 minutes. We chose 2 minutes as the time duration as it is sufficient to show differences in participants' performance; meanwhile they can keep focused.

We asked participants to make as few errors as possible - ensuring the success of precise placement as the primary requirement - and to go as far as they can - achieving better performance as the subordinate requirement. We designed this rule to prevent participants focusing only on making distance. They were asked to go back to the starting point each time they touch the metal path and trigger alarms.

5.3.3 Measurements

We aim to collect measurable data reflect user performance and user experience. Effectiveness and accuracy are measurements that indicate the performance of participants [26]. Effectiveness is measured by the distance achieved within a certain time. Meanwhile, accuracy can be quantified by errors committed within a certain time. Users' satisfaction and the cognitive load are used to indicate user experience. We used these three measurements (Distance, Errors, and Satisfaction level) for main statistical analysis. As the other measurements are either descriptive or indirect, thus they were mainly used for reason analysis. The overall collected measurements are:

1. Participants' performance
   - Distance
   - Errors (including numbers of overall errors, errors when using voice commands, and errors when being asked questions)
   - Uses amounts of voice commands to call for corresponded views (including the F/T/L/M views)
2. Participants' experience
   - Satisfaction level
   - Heartrate
   - Numbers of gaze shifts

During each test, we recorded the numbers of overall errors when the participants touched the metal path by counting the times of triggered alarms. We also marked the triggered errors when the participants using voice commands as well as the errors occurred during the interrupting questions in test C. Moreover, we recorded the amount of times when using voice commands to call for the offered views, including the real-time front view (F), the real-time top view (T), the real-time left view (L), and the static front view with a path map (M). Another measurement we collected during the tests was participants' gaze shifts between the two screens. We count one shift when a clear movement of the gaze focus were identified. We tracked this data by recording the participants' facial expression via cameras inbuilt on top of the screens. These measurements were collected for analyzing participants' cognitive state. We used a behavior observation form to record these measurements. Additionally, participants' heartrate data was recorded by an Android smartwatch and SenseIT app, which can be used to link the level of arousal.

After each test, we recorded the final distance. The final distance refers to the overall length achieved in one test. The distance participants achieved when making errors was eliminated to prevent participants ignoring the errors. An After-Scenario Questionnaire (ASQ) [77] was used immediately after each test to measure participants' satisfaction level. The ASQ is a brief 3-item questionnaire that measures components, such as ease of task completion, interactive modality to complete a task, and adequacy of support information, using a 7-step Likert response format (1 = strongly disagree, 7 = strongly agree).

After all tests, participants attended a short interview to give feedback to the general procedure. Multiple question formats were used to obtain the most effective results. The 7-step Likert response format was still used when asking participant's subjective opinions. Related materials could be found in the testing protocol (see Appendix III.)

5.3.4 Hypotheses

We used significant tests to analyze the difference of means. A paired-sample t-test method was adopted to compare different conditions in tests C and D with test B separately. Tests A and B were compared to identify differences between the indirect view and the direct view.

We formulated a one-tailed t-test hypothesis as tests C and D were conducted in a challenging situation in which the direct view was supposed to be less difficult. We predicted that both performance and experience in test B are better than in tests C and D. Meanwhile, we anticipated the result of test A to be more positive than that of test B. Specific hypotheses are as follows:

Comparison of test B with test C

Errors

$H0_{er.bc}$, there is no difference in the number of errors participants made between test B and test C.

If $H0_{er.bc}$ is rejected, then

$Ha1_{er.bc}$, participants make more errors in test B than in test C. ($Diff_{C-B}<0$)

$Ha2_{er.bc}$, participants make less errors in test B than in test C. ($Diff_{C-B}>0$)

Distance

$H0_{dif.bc}$, there is no difference in the distance participants achieved between test B and test C.

If $H0_{dif.bc}$ is rejected, then

$Ha1_{dif.bc}$, participants make more distance in test B than in test C. ($Diff_{C-B}<0$)

$Ha2_{dif.bc}$, participants make less distance in test B than in test C. ($Diff_{C-B}>0$)

Ease

$H0_{ea.bc}$, participants think there is no difference in ease of task completion between test B and test C.

If $H0_{ea.bc}$ is rejected, then

$Ha1_{ea.bc}$, participants think test B is harder than test C. ($Diff_{C-B}<0$)

$Ha2_{ea.bc}$, participants think test B is easier than test C. ($Diff_{C-B}>0$)

Satisfaction regarding voice interaction

$H0_{sa.v.bc}$, participants show no difference in satisfaction regarding voice interaction between test B and test C.

If $H0_{sa.v.bc}$ is rejected, then

$Ha1_{sa.v.bc}$, participants are more satisfied with test C than with test B. ($Diff_{C-B}<0$)

Ha2$_{sa.v.bc}$, participants are more satisfied with test B than with test C. ($Diff_{C-B}>0$)

Satisfaction regarding support information

H0$_{sa.i.bc}$, participants show no difference in satisfaction regarding support information between test B and test C.

If H0$_{sa.i.bc}$ is rejected, then

Ha1$_{sa.i.bc}$, participants are more satisfied with test C than with test B. ($Diff_{C-B}<0$)

Ha2$_{sa.i.bc}$, participants are more satisfied with test B than with test C. ($Diff_{C-B}>0$)

Comparison of test B with test D

Errors

H0$_{er.bd}$, there is no difference in the number of errors participants made between test B and test D.

If H0$_{er.bd}$ is rejected, then

Ha1$_{er.bd}$, participants make more errors in test B than in test D. ($Diff_{D-B}<0$)

Ha2$_{er.bd}$, participants make less errors in test B than in test D. ($Diff_{D-B}>0$)

Distance

H0$_{dif.bd}$, there is no difference in the distance participants achieved between test B and test D.

If H0$_{dif.bd}$ is rejected, then

Ha1$_{dif.bd}$, participants make more distance in test B than in test D. ($Diff_{D-B}<0$)

Ha2$_{dif.bd}$, participants make less distance in test B than in test D. ($Diff_{D-B}>0$)

Ease

H0$_{ea.bd}$, participants think there is no difference in ease of task completion between test B and test D.

If H0$_{ea.bd}$ is rejected, then

Ha1$_{ea.bd}$, participants think test B is harder than test D. ($Diff_{D-B}<0$)

Ha2$_{ea.bd}$, participants think test B is easier than test D. ($Diff_{D-B}>0$)

Satisfaction regarding voice interaction

H0$_{sa.v.bd}$, participants show no difference in satisfaction regarding voice interaction between test B and test D.

If H0$_{sa.v.bd}$ is rejected, then

Ha1$_{sa.v.bd}$, participants are more satisfied with test D than with test B. ($Diff_{D-B}<0$)

Ha2$_{sa.v.bd}$, participants are more satisfied with test B than with test D. ($Diff_{D-B}>0$)

Satisfaction regarding support information

H0$_{sa.i.bd}$, participants show no difference in satisfaction regarding support information between test B and test D.

If H0$_{sa.i.bd}$ is rejected, then

Ha1$_{sa.i.bd}$, participants are more satisfied with test D than with test B. ($Diff_{D-B}<0$)

$Ha2_{sa.i.bd}$, participants are more satisfied with test B than with test D. ($Diff_{D-B}>0$)

Comparison of test A with test B

Errors

$H0_{er.ab}$, there is no difference in the number of errors participants made between test A and test B.

If $H0_{er.ab}$ is rejected, then

$Ha1_{er.ab}$, participants make more errors in test A than in test B. ($Diff_{B-A}<0$)

$Ha2_{er.ab}$, participants make less errors in test A than in test B. ($Diff_{B-A}>0$)

Distance

$H0_{dif.ab}$, there is no difference in the distance participants achieved between test A and test B.

If $H0_{dif.ab}$ is rejected, then

$Ha1_{dif.ab}$, participants make more distance in test A than in test B. ($Diff_{B-A}<0$)

$Ha2_{dif.ab}$, participants make less distance in test A than in test B. ($Diff_{B-A}>0$)

Ease

$H0_{ea.ab}$, participants think there is no difference in ease of task completion between test A and test B.

If $H0_{ea.ab}$ is rejected, then

$Ha1_{ea.ab}$, participants think test A is harder than test B. ($Diff_{B-A}<0$)

$Ha2_{ea.ab}$, participants think test A is easier than test B. ($Diff_{B-A}>0$)

Satisfaction regarding support information

$H0_{sa.i.ab}$, participants show no difference in satisfaction regarding support information between test A and test B.

If $H0_{sa.i.ab}$ is rejected, then

$Ha1_{sa.i.ab}$, participants are more satisfied with test B than with test A. ($Diff_{B-A}<0$)

$Ha2_{sa.i.ab}$, participants are more satisfied with test A than with test B. ($Diff_{B-A}>0$)

5.4 Procedure

After the consent form of the study was signed, we provided participants training sessions to avoid learning effect. There were two training sessions in total before the real testing. The first one took place before the test A. Participants practiced with the spiral games to familiarize themselves with the rules. Another training section was set before the test B to help participants get used to giving voice command with the setup. We also mixed the order of tests C and D to eliminate potential inferences. The order of tests A and B were always maintained at the beginning as they are regarded as benchmark tests. As a result, half of the participants had test order ABCD, while the other half had test order ABDC.

5.5 Participants

Participants are students and employees at the University of Twente. One participant took all the tests to eliminate interference raised by individual variation. Thirteen participants participated in the study in 2 days. Their ages ranged from 19 to 65 (*M* = 29.67, *SD* = 13.28). Twelve participants are right-handed, while one is left-handed. Eleven participants reported being in a normal state, while one participant felt tired and another one was hungry while performing the tests. Gender distribution was balanced with 7 male and 6 females. Seven participants had test order ABCD, while the rest had test order ABDC. Test results of each participant are presented in Appendix IV.

5.6 Results

We presented the results and made a discussion with the order name of the test, we repeated the name and content of the four tests here to help the reading. The detail design of task could be found in Section 5.3.1.

Test A: Playing the game with d without the setup.
(Direct view / without voice commands)
Test B: Playing the game with the setup in the standard condition,
(Indirect view / with voice commands)
Test C: Playing the game with the setup when there were interrupting questions,
(Indirect view / with voice commands/ with interrupting questions)
Test D: Playing the game with the setup contains a more challenging route.
(Indirect view / with voice commands/ with challenging route)

5.6.1 Effects of interrupting question condition by comparing test B with C

A paired-samples t-test was adopted to determine if there was a difference when participants were asked interrupting questions while performing the test. We used an alpha level of .05 for all statistical tests and compared measurements of performance and experience, respectively. There was a statistically significant difference between test B (*M* = 5.85, *SD* = 1.21) and test C (*M* = 5.08, *SD* = 1.61, *t* (12) = -2,738, $p < .05$) regarding ease of task completion. As such, hypothesis $H0_{ea.bc}$ was rejected. Meanwhile, the one-tailed *p* value of $Diff_{C\text{-}B}<0$ is around 0.991 (P (Diff< 0) = 0.991). The alternative hypothesis $Ha1_{ea.bc}$ that participants think test B is harder than test C was, therefore, accepted. Notably, no significant difference in other measurements were noted for these two tests. Details of comparison between the other dependent variables could be found in Appendix IV Testing Results.

5.6.2 Effects of challenging route condition by comparing test B with D

A paired-samples t-test was used to determine if there was a difference when participants performed a more challenging route. Results reveal a statistically significant difference in error between test B (*M* = 5.62, *SD* = 1.50) and test D (M = 4.62, SD = 1.45, t (12) = 2.944, p < .05). Accordingly, hypothesis $H0_{er.bd}$ was rejected. The one-tailed *p* value of $Diff_{D\text{-}B}<0$ is around 0.994 (P (Diff< 0) = 0.994). For this reason, the alternative hypothesis $Ha1_{er.bd}$ that more errors are made in test B than in test D was accepted. There was no significant difference in other measurements for these two tests (see Appendix IV).

5.6.3 Effects of indirect view by comparing test A with B

We also compared results between test A and test B (see Appendix IV) with the alpha level of .05. Results indicate a significant difference in errors between test A (*M* = 4.15, *SD* = 2.15) and test B (*M* = 5.62, *SD* = 1.50) using a two-tailed significant t-test (*t* (12) = -2.260, *p* =0.043). As such, we can reject $H0_{er.ab}$. The one-tailed *P* (Diff< 0) = 0.022 and *P* (Diff>0) = 0.978; therefore, we considered the alternative hypothesis $Ha2_{er.ab}$ that less errors are made in test A than in test B. We then compared the overall distance. Results indicate a significant difference in distances between test A (*M* = 150.10, *SD* = 98.71) and test B (*M* = 9.92, *SD* = 3.79), *t* (12) = 5.168, *p* = 0.0002). As such, $H0_{dis.ab}$ was rejected as well. The possibility of Diff<0 is 0.999; therefore, $Ha1_{dif.ab}$ was accepted. These findings affirm that participants make more distance in test A than in test B.

Moreover, we compared user experience between test A and test B based on ease of task completion and satisfaction regarding support information (see Appendix IV). The two-tailed *p* value at the confidence level .05 is less than 0.05, which indicates a significant difference between test A (*M* = 2,92, *SD* = 1.44) and test B (*M* = 5.85, *SD* = 1.21) regarding ease of task completion (*t* (12) = -8.875, *p* = $1.3*10^{-6}$.) The one-tailed *P* (Diff>0) is even less than $6.4*10^{-7}$; therefore, the level of ease in test A is lower than that of test B. User satisfaction regarding support information also shows a significant difference between test A (*M* = 1.92, *SD* = 1.32) and test B (*M* = 3.69, *SD* = 1.55, *t* (12) = -3.320, *p* = 0.006). The one-tailed *P* (Diff< 0) = 0.003 and *P* (Diff>0) = 0.997 indicate higher satisfaction level in test A than in test B; therefore, $Ha2_{sa.i.ab}$ was accepted.

5.6.4 Other results

In this section, we compared the mean, the standard deviation, and the distribution of other measurements taken during the tests for further analysis. These data were collected to explain the results above. We collected data on gaze shifts (see Table 4), which are most frequent in test B (*M* =25.31, *SD* =11.45), followed by test D (*M* = 24.69, *SD* = 15.01). The lowest average times of gaze shifts are in test C (*M* = 22.08, *SD* =9.19). However, after analyzing significance, results reveal no significant difference between test B and test C (*t* (12) = 1.592, *p* = 0.067). The *P* value obtained comparing test B and test D is even higher (*t* (12) = 0.231, *p* = 0.411).

We also manually recorded the errors participants made when they were asked interrupting questions in test C (see Table 5). Results indicate that participants made 1.62 errors (*M* = 1.62, *SD* = 1.76) when they were asked a question. They made 5.85 errors in total (*M* = 5.85, *SD* = 2.38). In addition, we manually recorded the errors participants made when they were using a voice command (see Table 6). Average errors when using voice commands were the same in test B (*M* = 1.15, *SD* = 0.99) and test D (*M* = 1.15, *SD* = 1.14). Less errors were made in test C (*M* = 1.00, *SD* = 1.00). This type of errors accounts for 20.5% of overall errors in test B, 19.9% in test C, 24.9% in test D, and 22.10% in all tests.

Table 4: Overview of the number of gaze shifts, the means, and the standard deviation in each test.

	Test B	Test C	Test D
Gaze shift	8	13	26
	8	12	6
	29	16	19
	35	28	35
	12	12	12
	29	21	17
	36	22	53
	43	37	50
	29	38	24
	25	27	21
	34	30	35
	13	14	5
	28	17	18
M	25,31	22,08	24,69
SD	11,45	9,19	15,01

Table 5: Manual record of errors participants made when they were asked questions.

	Error when asking questions													
P0	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	M	SD
2	2	0	6	0	1	0	1	4	2	2	1	0	1.62	1.76

Table 6: Manual record of errors participants made when they used voice commands.

	Test B		Test C		Test D	
	Errors when using voice command	Overall errors	Errors when using voice command	Overall errors	Errors when using voice command	Overall errors
	0	5	3	6	2	4
	0	6	0	7	1	4
	2	7	0	3	0	5
	3	7	0	7	1	5
	0	4	0	5	0	4
	1	8	2	6	0	8
	2	4	2	3	0	2
	1	5	1	6	2	4
	1	7	0	3	3	6
	1	3	2	7	1	4
	2	7	1	3	0	4
	2	5	1	5	2	6
	0	5	1	5	3	4
M	1,15	5,62	1	5,08	1,15	4,62
SD	0,99	1,50	1,00	1,61	1,14	1,45

Participants' subjective opinions were also collected for analysis. They were asked the question "What do you think is the main reason you triggered the alarm?" and were provided a set of choices. Nine participants chose the reason "My arms are so tired of holding the hand shank for a long time." Meanwhile, seven participants answered, "There are too many views, and I cannot take all of them into account." Six participants selected the explanation "After I gave out voice commands, the view I needed was not shown to me in time." Meanwhile, four participants chose the reason "I get distracted when using voice to adjust view"; three claimed "The device is too sensitive to trigger an alarm"; two thought "The people surrounding me attract my attention"; and two answered, "I just want to finish the game as fast as possible, I don't care about the alarm." Only one participant pointed out "I get used to watching my hand, so I move my attention out of the screen." Participants also provided other reasons. One participant stated, "The view is not very clear for me; it seems that the front view makes great sense for operation." Another participant complained, "The isolated tape is slippery; the top camera has a delay." Moreover, participants provided customized explanations, such as "Game is too complex; the route is too curved"; "the front view disappear too soon"; and "the offered views are missing depth." One participant remarked, "I am too hungry, so I can't focus."

Based on the Likert scale (1 = strongly agree, 7 = strongly disagree), participants showed higher disagreement with the statement "I frequently switch my attention among different views" (M = 4.46, SD = 1.66) than the statement "I only focus on one view" (M = 3.46, SD = 1.90). Notably, participants demonstrated the highest disagreement score for the statement "I can easily answer the question from the host without interrupting the game" (M = 4.62, SD = 2.20). For the statement "I know which view I needed during the game and give out the right voice commands," participants had an average agreement score of 2.88 with standard deviation at 1.50 (M = 2.28, SD = 1.50). Meanwhile, the average for the statement "I think the four views offered are sufficient for completing the task" was 3.85 (M = 3.85, SD = 1.99). Participants exhibited the most agreement with the statement, "I fully focused on the game and didn't notice the surroundings" (M = 2.23, SD = 1.54).

At the end of the questionnaire, participants provided answers to the question "Which task do you think is the most difficult?" Eight participants chose test C, while three chose test D. Two participants chose test B and explained, "Test B comes first, and it is a sudden increment of difficulties. I thought test B is the hardest. But after finishing the following two tasks, I know it could be even harder." Participants had a consistent opinion regarding the question "Which view do you think is the most helpful?" All of them chose the real-time front view (F).

5.7 Discussion

5.7.1 Reasons for the opposing results.

Data collected from the study did not correspond to the hypotheses. Some even exhibited conflicting trends. Only two significant differences were noted when the standard condition was compared with the challenging conditions. For ease of task completion, tests B and C demonstrated statistical difference. Significant difference was observed in errors participants made by comparing tests B to D. In contrast to our prediction, the following alternative hypotheses were accepted:

Participants thought test B was harder than the test C and participants made more errors in test B than in test D.

Reasons for these opposing trends are varied. First, learning effect influences the results of the tests. Although we offered a pre-training session before testing, we changed the order of tests B and C to eliminate learning effect. However, test B (i.e., standard task) was always conducted before tests C and D. Results reveal that participants achieved better performances in the following tasks than in earlier tasks. Second, although the pre-training helps familiarize participants with the rules and setup, they were more focused on and emulous in the official task. Participants learned to achieve better results under time pressure. This is also reflected in the ASQ. Participants showed most agreement regarding the statement “I fully focused on the game and didn’t notice the surroundings” (*M* = 2.23, *SD* = 1.54). Two participants admitted “[they] just want to finish the game as fast as possible and [didn’t] care about the alarm” even though they were asked to make as few errors as possible.

5.7.2 Reasons for conflicting results after sub-tests and after all tests.

Participants provided conflicting feedback regarding the level of ease for task completion at different timing. Based on the ASQ results obtained immediately after participants completed each task, the highest mean score is 5.85 in test B, followed by 5.27 in test D, and 5.08 in test C. Ten participants gave a score of 6 or above to test B; seven gave this score to test D; and six participants in test C gave the same level of scores. However, when participants finished all the tests, they had conflicting evaluation regarding task difficulty were collected. When asked “Which task do you think is the most difficult?”, eight participants voted for test C; three voted for test D; and only two chose test B.

This above evaluation may be because, compared to test A, the difficulty of test B suddenly increased and overwhelmed the participants. Results in Section 5.6.3 indicate a significant difference in errors, distance, ease, and satisfaction between tests A and B, especially the distance participants achieved. The mean values in test A (*M* = 150.10, *SD* = 98.71) is 15 times higher than that in test B (*M* = 9.92, *SD* = 3.79). Participants were not prepared for the upcoming difficulties, so they gave high scores after completing tasks in test B. When performing the tasks in tests C and D, they were more mentally prepared. The gap between their expectation and reality, in turn, were modified. One participant explained his opposing feedback when selecting the most difficult condition, “I thought condition B is the hardest, but after doing the other tasks, I found it could be even harder.” Since test B came first, it caused a sudden increment in difficulty.

5.7.3 Cognitive state of participants

Results of the comparison between test A and test B reveal that both user performance and user experience were lower when they were performing tasks with the setup. Section 5.6.3 mentioned a significant difference in errors, distance, ease, and satisfaction between tests A and B, especially the distance participants achieved. These values also indicate high cognitive load when performing the task with an indirect view.

Based on participants' feedback, we observed limited cognitive resource when taking all the views into account. According to seven participants, the reason the alarm was triggered is "There are too many views, and I cannot take all of them into account.". More participants also agree with the statement "I only focus on one view" (*M* = 3.46, *SD* = 1.90) than the statement "I frequently switch my attention among different views" (*M* = 4.46, *SD* = 1.66).

Participants experienced more cognitive challenge when someone asked them questions during the task. Participants made in average 1.62 errors (*SD = 1.76*) when asked a question which accounted for 27.2% total errors (see Section 5.6.4). Participants showed the highest disagreement score for the statement "I can easily answer the question from the host without interrupting the game" (*M* = 4.62, *SD* = 2.20). Notably, participants had the least gaze shifts in test C (*M* = 22.08, *SD* = 9.19), indicating that they have even less capacity to take four views into account when they are asked questions (see Section 5.6.4)

Cognitive overload was also identified in the manual record of errors that participants made when using voice commands (see 5.6.4). Average errors when using voice commands were almost identical in three tests: test B (*M* = 1.15, *SD* = 0.99), test C (*M* = 1.00, *SD* = 1.00), and test D (*M* = 1.15, *SD* = 1.14)]. This type of errors account for 22.1% total errors. Four participants answered, "I get distracted when using voice to adjust views."

To summarize, we did not observe significant differences in user performance and user experience amid challenging conditions. Results, however, indicate high cognitive load among participants performing precise placement with the indirect view. Moreover, voice commands induce potential cognitive overload among participants. A similar effect is possible when radiologists use voice commands in the context of IR, especially when there is random conversation.

5.7.4 Reasons of the insufficient results

Most of the results did not indicate significant differences. A possible reason for this may be that the samples we collected were insufficient. Thirteen samples in a simulated test is not enough. Another reason may be that the indirect view is challenging for participants, so the result did not show a significant difference even when all three conditions require full capacity. In the ASQ, participants rated all three conditions as "hard"; in fact, for ease of task completion, the average score was over 5. Some participants provided the following reasons alarms were triggered: "The device is too sensitive to trigger an alarm". "Game is too complex; the route is too curved"; and "Front view disappeared too soon." These reasons indicate that task completion with the in-direct view is challenging.

Another invalid data was the collected heartrate. Because of technical problems, the collected heartrate couldn't be used to map the timeline of each subtests. Therefore, we didn't use them to interpret the results.

5.7.5 Limitations of this study

The present study has several limitations, particularly when running the tests and collecting data.

1. Some data were insufficient and invalid as we explained above.
2. Learning effect influences test results. Participants perform better in later conditions, and therefore give a more optimistic grade.
3. The setup is not friendly for left-handed participants. They can still perform the task, but flexibility is worse than for right-handed participants because of the location of the hole.
4. There is a delay in showing the corresponding view after participants have given out voice commands, which in turn influenced their performance.
5. The way we calculated distance can be optimized as we only recorded the overall length achieved and eliminated the distance participants achieved when they made errors. Although we decided to follow this rule to prevent participants from ignoring errors, we can still add the distance they made before making errors to the results. Participants would then be asked go back to the starting point.
6. Participants have different traits—some are very focused during the tests, but when they are hungry or tired, they get distracted easily. The performance of participants in different physical conditions should be recorded separately.

Several improvements are recommended for future work. More comprehensive training on performing the task under time pressure is needed. More participants should be encouraged to participate in the study, and their performance should be recorded separately. Participants who are left-handed should be considered as well. A functional VI system can be used for evaluation. Methods to collect the overall distance and heart rate can also be improved.

5.8 Conclusion

In this study, we examined how challenging conditions influence the performance and satisfaction of users in a simulated context where simple voice commands were offered to display required views. Playing the spiral game in a box via cameras was used to mimic the way radiologists conduct IR. We adopted controlled variable methods to run a within-subject study. The independent variables were the different test conditions, including the direct view condition (test A), the standard condition (test B), the condition with interrupting questions (test C), and the condition with the challenging route (test D).

Observation forms and the ASQ were used to collect data that measure user performance and experience in the four tests. We also used other data, such as times of voice commands, gaze shifts, and other related information, for further analysis. Moreover, an after-test interview was conducted to collect feedback from participant.

Because of multiple limitations, we did not achieve significant differences in the performance and user experience when challenging conditions occurs (tests C and D). But the results presented a significant difference between without (test A) and with the set-ups (test B). From this, we interpreted a high cognitive load of precise placement via an in-direct view. Participants also reported difficulties of taking all views into account. Another finding is voice commands may bring potential cognitive overload. Evidence can be found from the error’s participants made when they are giving voice commands. Distractions could also be identified when an interrupting question appeared. More proves could be uncovered in participants’ subjective feedbacks. Thus, we made a conclusion: VI may bring potential cognitive overload in IR where radiologists’ mental consumption is comparatively high, especially when a random conversation occurs. Chapter 6 provides the final conclusion of this study, clarifies its value and contributions, and outlines suggestions for future research.

6 CONCLUSION

6.1 Summary

In this thesis study, we explored VI solutions in IR, as the hands of interventional radiologists are often occupied, holding wires and catheters. This situation thus limits the radiologists' autonomy and their efficiency in completing tasks. Rapid development in VI technology can potentially address this issue.

To explore VI solutions for the hands-occupied problem, we first conducted background research to understand the features, the current problems, and related studies in IR. Then we focused on the hands-occupied problem. We elaborated the user's condition and detailed interaction by interviewing the related experts and watching video records. However, the collected results from literature and context analysis were still descriptive and complex. So, we adopted a synthesize stage to disintegrate the problem clearer. Four conditions of the hands-occupied problem were proposed. We also explained the user requirements for further solution exploration. A set of personas and a scenario were then used to construct a more concrete and contextual interactive story. After analyzing the interactions in this scenario, five design opportunities and two challenges were discussed. Then we introduced a study with four sub tests to ascertain the effect of introducing VI into IR, especially when a challenging condition appears. A simulated setup was constructed. Results related to user performance and experience were collated to answer the research question.

6.2 Conclusion

The main research question of this research is as follows:
Can VI technology solve the hands-occupied problem in IR and further improve interventional radiologists' performance and experience?

After this study, we concluded that: It's hazardous to making use of VI in IR, where radiologists' mental consumption is comparatively high. VI may bring potential cognitive overload, especially when a random conversation occurs. We argue a cautious and well-considered usage of speech technology in IR. Detailed arguments are presented as following.

The research question is divided into two sub-questions, and gradually answered in this research. First, can VI solve the hands-occupied problem? Second, can the solution also benefit interventional radiologists through better performances and user experience? To answer the first question (i.e., Can VI solve the hands-occupied problem?), opportunities in the IR procedure were identified. After analyzing users' state and detailed interactions, five tasks were identified as opportunities for using VI to meet user requirements (e.g., independence).

- Move the C-arm along with the inserted wires and catheters

- Set/adjust parameter of contrast media with the inject pump
- Ask for display of captured images during the diagnosis
- Move the C-arm to a standard position when diagnosis disease
- Slightly adjust the angle of C-arm to recognize the path of arteries

However, it is challenging to solve the hands-occupied problem and improve user performance and user experience at the same time. Two challenging conditions were proposed. The first one was, when the primary task is complicate, for example, an unusual patient anatomy appeared. Another challenging condition was, when someone asks them questions. We tried to answer the second sub-questions ((i.e., Can VI both benefit performance and experience?) by conducting a study. A simulated context was used to test how challenging conditions influence the performances and experience of users when voice commands are used.

Although no significant difference regarding user performance and user experience can be concluded between challenging conditions and standard conditions, results indicated high cognitive load among participants performing precise placement with the indirect view. A significant difference both in performance and experience was identified between without (test A) and with the set-ups (test B). Participants also reported difficulties of taking all views into account. Voice commands may bring potential cognitive overload. Proof can be found from the error's participants made when they giving voice commands (22.10% in all errors). Distractions by an interrupting question were found. 27.2% of total errors occurred when participants being asked questions. More proves were founded in participants' subjective feedbacks (see Section 5.7.3）. These results denoted potential of cognitive overload in IR when radiologists using voice commands to interact with the system, especially when there is a random conversation.

Therefore, we [do not] recommend introducing VI in IR without a well-considered manner. The cognitive state of radiologists and complexity of IR are two critical factors. This research has some limitations.

6.3 Discussion

First, there is a difference between the simulated context and real context. The simulated task in the study may be harder than the real one. There are differences between the way of playing the spiral game and conducting IR. During the tests, the alarm in the spiral game was easy to trigger. Operating the hook under the camera was more challenging than manipulating wires and catheters in IR. The precise placement should be easier after extensive practice. We didn't have the opportunity to observe a real IR treatment. Therefore, the simulation is based on the features we identified from the interviews and video observations. It may also limit the results. Thus, we strongly recommend future work run a test in real context with real radiologists.

Second, the participants' cognitive state was measured indirectly. We adopted indicators to draw a conclusion, such as gaze shifts, errors when using voice commands and subjective feedback of the participants. In this study, the cognitive state of radiologists was highlighted both in theoretical analysis and study results. We propose future research to evaluate radiologist's' cognitive load more directly. Solutions that can decrease the cognitive load radiologists are imperative. If future researchers keep working on a touchless solution, we would recommend a focus on modality fusion to ease radiologists' work.

This research was conducted within a limited scope. The results can be used as a reference when researchers work on other IR procedures or techniques.

6.4 Future work

If other research groups or we could work on the topic iteratively or further, we would like to suggest some improvement tips.

First of all, as we mentioned before, we were limited with the resource accessibility. We strongly recommended running field research, observing a real and complete IR procedure, interviewing IR experts at their worksite, asking for explanation or demonstration of the working system, and may as well evaluate a working prototype in a real hybrid operating room. The effect of results gathered from a simulated context was limited. In other words, if the simulation was not defined rigorously, it is possible to achieve different results in the real context.

Another tip was, it would be better to start the research with a limited scope. Though we understood the background of IR more comprehensively, it also made the problem more complicated and set the goal vaguer. A focus at the beginning would be much efficient and productive. I would suggest choosing one specific task, then explore targeted solutions with considering the IR context and the user's state.

If we could improve this study according to the two tips mentioned above, we would further suggest a comparison study to ascertain the difference between relying on an assistant and using a voice interaction system. We cannot run such comparative testing, as simulating two roles (radiologists and assistants) was challenging, especially when they sharing a certain common knowledge. This kind of comparison is more direct and competent to decide whether introducing VI to IR. However, the real users and using context were vital.

At last, we also proposed several other interesting directions for future studies. Firstly, as the cognitive state of radiologists would be a challenge when introducing new technologies, further studies that attempt to explain cognitive processes are also significant. Secondly, future research can study problems other than the hands-occupied problem that interventional radiologists encounter (see Section 2.4.2). More professional insights are needed to develop a user-friendly solution. Thirdly, future studies can research on other IR technique, such as tumor intervention or neuro-intervention. These techniques are conducted with different setups and team composition. Last but not least, this research focused on a VI solution to address the hands-occupied problem. Future work can explore other kinds of touchless solutions.

REFERENCE

[1] J. H. Tan, C. Chao, M. Zawaideh, A. C. Roberts, and T. B. Kinney, “Informatics in Radiology: Developing a Touchless User Interface for Intraoperative Image Control during Interventional Radiology Procedures,” *RadioGraphics*, vol. 33, no. 2, pp. E61–E70, 2012.

[2] A. Mahnken, A. Bücker, C. Hohl, and A. Berlis, “White Paper: Curriculum in Interventional Radiology,” *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der Bildgeb. Verfahren*, vol. 189, no. 04, pp. 309–311, Mar. 2017.

[3] M. Pfister *et al.*, “Real-time respiratory motion tracking: roadmap correction for hepatic artery catheterizations,” in *Medical Imaging 2008: Visualization, Image-guided Procedures, and Modeling*, 2008.

[4] G. Nollert *et al.*, “The Hybrid Operating Room,” in *Special Topics in Cardiac Surgery*, 2007.

[5] D. L. Miller, “OVERVIEW OF CONTEMPORARY INTERVENTIONAL FLUOROSCOPY PROCEDURES,” *Health Phys.*, vol. 95, no. 5, pp. 638–644, Nov. 2008.

[6] American society of radiologic technologists, “Radiologic Sciences Staffing and Workplace Survey,” 2015.

[7] A. Mewes, B. Hensen, F. Wacker, and C. Hansen, “Touchless interaction with software in interventional radiology and surgery: a systematic literature review,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 2, pp. 291–305, 2017.

[8] J. A. Brink, R. L. Arenson, T. M. Grist, J. S. Lewin, and D. Enzmann, “Bits and bytes: the future of radiology lies in informatics and information technology,” *Eur. Radiol.*, vol. 27, no. 9, pp. 3647–3651, 2017.

[9] I. RADO, *The positioning of the c-arm and the patient during the minimally invasive pain treatment*, vol. 113, no. 2. 2011.

[10] A. Hübler, C. Hansen, O. Beuing, M. Skalej, and B. Preim, “Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining,” Otto-von-Guericke University Magdeburg.

[11] R. Johnson, K. O’Hara, A. Sellen, C. Cousins, and A. Criminisi, “Exploring the potential for touchless interaction in image-guided interventional radiology,” in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI ’11*, 2011, p. 3323.

[12] C. D. Williams and B. Short, “ACR and ASRT development of the radiologist assistant: Concept, roles, and responsibilities,” *J. Am. Coll. Radiol.*, vol. 1, no. 6, pp. 392–397, 2004.

[13] A. Iannessi, P. Y. Marcy, O. Clatz, P. Fillard, and N. Ayache, “Touchless intra-operative display for interventional radiologist,” *Diagn. Interv. Imaging*, vol. 95, no. 3, pp. 333–337, 2014.

[14] R. Article, “A REVIEW ON APPROACHES TO DEVELOP GESTURE AND

VOICE RECOGNITION TECHNIQUE," vol. 2, no. 3, pp. 86–88, 2013.

[15] C. A. O. Nathan, V. Chakradeo, K. Malhotra, H. D'Agostino, and R. Patwardhan, "The voice-controlled robotic assist scope holder AESOP for the endoscopic approach to the sella," *Skull Base*, vol. 16, no. 3, pp. 123–132, 2006.

[16] P. R. Cohen and S. L. Oviatt, "The role of voice input for human-machine communication.," *Proc. Natl. Acad. Sci.*, vol. 92, no. 22, pp. 9921–9927, 2006.

[17] C. Kamm, "User interfaces for voice applications.," *Proc. Natl. Acad. Sci.*, vol. 92, no. 22, pp. 10031–10037, Oct. 1995.

[18] L. H. Schwartz, P. Kijewski, H. Hertogen, P. S. Roossin, and R. A. Castellino, "Voice recognition in radiology reporting.," *Am. J. Roentgenol.*, vol. 169, no. 1, pp. 27–29, 1997.

[19] G. Nollert and S. Wich, "Planning a Cardiovascular Hybrid Operating Room: the Technical Point of View," *Heart Surg. Forum*, vol. 12, no. 3, pp. E125–E130, Jun. 2009.

[20] M. Peissner, V. Doebler, and F. Metze, "Can voice interaction help reducing the level of distraction and prevent accidents?," *Meta-Study Driv. Distraction Voice Interact. White Pap.*, no. May, p. 24, 2011.

[21] M. B. Hoy, "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants," *Med. Ref. Serv. Q.*, 2018.

[22] L. A. Schwarz, A. Bigdelou, and N. Navab, "Learning gestures for customizable human-computer interaction in the operating room," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011.

[23] A. Bigdelou, L. Schwarz, and N. Navab, "An adaptive solution for intra-operative gesture-based human-machine interaction," in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces - IUI '12*, 2012, p. 75.

[24] H. M. Mentis *et al.*, "Voice or Gesture in the Operating Room," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, 2015, pp. 773–780.

[25] J. Zimmerman, S. Evenson, and J. Forlizzi, "Discovering knowledge in the design case," *Proc. Futur.*, vol. 04, no. November, 2004.

[26] J. L. Jinjuan, H. Feng, and H. Hochheiser, "Research Methods in Human-Computer Interaction Second Edition," *J. Am. Soc. Inf. Sci. Technol.*, 2017.

[27] J Vasc Interv Radiol 2010;, "Global Statement Defining Interventional Radiology," pp. 1147–1149, 2010.

[28] J. Rösch, F. S. Keller, and J. A. Kaufman, "The birth, early years, and future of interventional radiology," *J. Vasc. Interv. Radiol.*, vol. 14, no. 7, pp. 841–853, 2003.

[29] D. A. Bluemke and S. Liu, "Imaging in Clinical Trials," in *Principles and Practice of Clinical Research (Third Edition)*, 2012, pp. 579–617.

[30] S. O. Hansson, "Ethics and radiation protection," *J. Radiol. Prot.*, vol. 27, no. 2, pp. 147–156, Jun. 2007.

[31] C. C. Blackmore, "Fundamentals of clinical research for radiologists. Data collection in radiology research," *Am. J. Roentgenol.*, vol. 177, no. 4, pp. 755–761, 2001.

[32] X. Chen and S. T. C. Wong, *Cancer Theranostics*. Elsevier, 2014.

[33] J. Anderson *et al.*, "Virtual reality training in interventional radiology: The Johns Hopkins and Kent Ridge digital laboratory experience," *Semin. Intervent. Radiol.*, vol. 19, no. 2, pp. 179–185, 2002.

[34] American Society of Radiologic Technologists, "Radiation Therapy , Management and Dosimetry Workplace Survey 2010," 2010.

[35] American society of radiologic technologists, "Radiologic Sciences Staffing and Workplace Survey 2017," 2017.

[36] D. Siddoway, M. Lou Ingeholm, O. Burgert, T. Neumuth, V. Watson, and K. Cleary, "Workflow in interventional radiology: nerve blocks and facet blocks," in *Medical Imaging 2006: PACS and Imaging Informatics*, 2006, vol. 6145, p. 61450B.

[37] D. Kessel and L. Robertson, *interventional radiology a survivial guide*. 2017.

[38] H. Humphreys *et al.*, "Guidelines on the facilities required for minor surgical procedures and minimal access interventions," *Journal of Hospital Infection*. 2012.

[39] T. Weirick and H. V. Anderson, "Percutaneous coronary intervention versus coronary artery bypass grafting for severe coronary artery disease," *Curr. Cardiovasc. Risk Rep.*, 2009.

[40] J. Kpodonu, "Hybrid cardiovascular suite: The operating room of the future," *Journal of Cardiac Surgery*, vol. 25, no. 6. pp. 704–709, 2010.

[41] M. Friebe, "Computed tomography and magnetic resonance imaging contrast media injectors : technical feature review – what is really needed ?," pp. 231–239, 2016.

[42] P. Kamusella, C. Wissgott, P. Wiggermann, F. Scheer, and R. Andresen, "Reduction of radiation exposure for the examiner in angiography using a direct dosimeter," *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgeb. Verfahren*, 2013.

[43] I. A. E. Agency, "Applying radiation safety standards in diagnostic radiology and interventional procedures using X rays," vol. 39, p. 110, 2006.

[44] M. N. Krishnan, "Peripheral angiography: Views and technique," *J. Indian Coll. Cardiol.*, vol. 6, pp. 114–117, 2016.

[45] I. Nisky, F. Huang, A. Milstein, C. M. Pugh, F. A. Mussa-Ivaldi, and A. Karniel, "Perception of stiffness in laparoscopy - The fulcrum effect," *Stud. Health Technol. Inform.*, vol. 173, pp. 313–319, 2012.

[46] M. Herrlich *et al.*, "Instrument-mounted displays for reducing cognitive load during surgical navigation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 9, pp. 1599–1605, 2017.

[47] A. Johnstone and G. Altmann, "Automated speech recognition," vol. 12, no. 1, pp. 239–243, 2007.

[48] R. O. N. King and L. Researcher, "Speech and Voice Recognition White Paper," 2014.

[49] K. B. Lee and R. A. Grice, “The design and development of user interfaces for voice application in mobile devices,” *IEEE Int. Prof. Commun. Conf.*, pp. 308–320, 2006.
[50] M. Johnson *et al.*, “A systematic review of speech recognition technology in health care,” *BMC Med. Inform. Decis. Mak.*, vol. 14, no. 1, 2014.
[51] P. Cohen and S. Oviatt, “The Role of Voice in Human-Machine Communication,” in *Voice Communication Between Humans and Machines*, 2015.
[52] L. H. Schwartz, P. Kijewski, H. Hertogen, P. S. Roossin, and R. A. Castellino, “Voice recognition in radiology reporting,” *Am. J. Roentgenol.*, vol. 169, no. 1, pp. 27–29, 1997.
[53] J. Lumsden, K. Klinger, J. Langel, K. Smalley, and A. Thor, “Handbook of research on user interface design and evaluation for mobile technology,” *Choice Rev. Online*, vol. 45, no. 12, pp. 45-6836-45–6836, 2013.
[54] H. JD and R. FW, “Experience with implementation of a radiology speech recognition system.,” *J. Digit. Imaging*, vol. 13, no. 3, pp. 124–128, 2000.
[55] J. A. Pezzullo, G. A. Tung, J. M. Rogg, L. M. Davis, J. M. Brody, and W. W. Mayo-Smith, “Voice recognition dictation: Radiologist as transcriptionist,” *J. Digit. Imaging*, vol. 21, no. 4, pp. 384–389, 2008.
[56] J. du Toit, R. Hattingh, and R. Pitcher, “The accuracy of radiology speech recognition reports in a multilingual South African teaching hospital,” *BMC Med. Imaging*, vol. 15, no. 1, pp. 1–5, 2015.
[57] R. Parente, N. Kock, and J. Sonsini, “An analysis of the implementation and impact of speech-recognition technology in the healthcare sector.,” 2004.
[58] F. Allison, M. Carter, and M. Gibbs, “A History of Voice Interaction in Games,” *Proc. 1st Int. Jt. Conf. DiGRA FDG*, pp. 1–2, 2016.
[59] Y. Miki, N. Yoshitsugu, and K. Ito, “Driver’s load of voice interaction system in vehicles,” pp. 1–6, 2003.
[60] J. Migneault, J. M. Robert, M. C. Desmarais, and S. Caron, “Comparing voice with touch screen for controlling the instructor’s operating station of a flight simulator,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, pp. 3662–3667, 2007.
[61] A. D. Baddeley and G. J. Hitch, “Developments in the Concept of Working Memory,” *Neuropsychology*, 1994.
[62] G. Wernaart, “Cognitive Load Measurement : Different instruments for different types of load ?,” p. 30, 2012.
[63] K. L. Fraser, P. Ayres, and J. Sweller, “Cognitive load theory for the design of medical simulations,” *Simul. Healthc.*, vol. 10, no. 5, pp. 295–307, 2015.
[64] C. L. Sistrom, “Conceptual approach for the design of radiology reporting interfaces: The talking template,” *J. Digit. Imaging*, vol. 18, no. 3, pp. 176–187, 2005.
[65] Y. Cao, M. Theune, and A. Nijholt, “Towards cognitive-aware multimodal presentation: The modality effects in high-load HCI,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5639 LNAI, pp. 3–12, 2009.

[66] A. M. Hötker, M. B. Pitton, P. Mildenberger, and C. Düber, “Speech and motion control for interventional radiology: Requirements and feasibility,” *Int. J. Comput. Assist. Radiol. Surg.*, 2013.
[67] R. Wipfli, V. Dubois-Ferrière, S. Budry, P. Hoffmeyer, and C. Lovis, “Gesture-Controlled Image Management for Operating Room: A Randomized Crossover Study to Compare Interaction Using Gestures, Mouse, and Third Person Relaying,” *PLoS One*, vol. 11, no. 4, p. e0153596, 2016.
[68] L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross, “You can’t touch this: Touch-free navigation through radiological images,” *Surg. Innov.*, 2012.
[69] M. Ye, E. Johns, B. Walter, A. Meining, and G. Z. Yang, “An image retrieval framework for real-time endoscopic image retargeting,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 8, pp. 1281–1292, 2017.
[70] G. E. H. El-Shallaly, B. Mohammed, M. S. Muhtaseb, A. H. Hamouda, and A. H. M. Nassar, “Voice recognition interfaces (VRI) optimize the utilization of theatre staff and time during laparoscopic cholecystectomy,” *Minim. Invasive Ther. Allied Technol.*, 2005.
[71] I. A. Salama and S. D. Schwaitzberg, “Utility of a Voice-Activated System in Minimally Invasive Surgery,” *J. Laparoendosc. Adv. Surg. Tech.*, 2005.
[72] A. Beristain, A. De Mauro, K. Van De Weyer, and D. Segers, “Voice controller for Image Guided Surgery and personalized Interactive Visualization,” *Proc. Iwbbio 2014 Int. Work. Bioinforma. Biomed. Eng. Vols 1 2*, pp. 1244–1253, 2014.
[73] A. Alapetite, “Speech recognition for the anaesthesia record during crisis scenarios,” *Int. J. Med. Inform.*, 2008.
[74] S. A. Jacob and S. P. Furgerson, “Writing interview protocols and conducting interviews,” *Virus Res.*, vol. 21, no. 3, pp. 199–212, 1991.
[75] N. Lavie, “Perceptual Load as a Necessary Condition for Selective Attention,” *J. Exp. Psychol. Hum. Percept. Perform.*, 1995.
[76] N. Lavie, “Distracted and confused?: Selective attention under load,” *Trends in Cognitive Sciences*. 2005.
[77] J. R. Lewis, “Psychometric evaluation of an after-scenario questionnaire for computer usability studies,” *ACM SIGCHI Bull.*, 2007.
[78] G. Faragò, V. Caldiera, G. Tempra, and E. Ciceri, “Advanced digital subtraction angiography and MR fusion imaging protocol applied to accurate placement of flow diverter device,” *Journal of NeuroInterventional Surgery*. 2016.

1 APPENDIX INTERVIEW PROTOCOL

Pre-testing consent form

Goal:
This interview is to gain expert opinions about procedures and teamwork of the interventional radiology, especially the angiography procedure. This interview is contributing to the user research for the master thesis voice interaction within radiology department. Commercial usage are forbidden.

Procedure:
The interview will be basically in the form of question and answer. There are three blocks in totally, and the whole procedure will be recorded.

Confidentiality:
According to the confidential contract (NDA), all the information gained from this interview are only used for this master thesis research. The Name of the interviewee will not be shown. And the research result won't be published in 5 years.

Alternative to Participation:
Participation in this study is voluntary. You are free to withdraw or discontinue participation at any time.

Cost and Compensation:
Participation in this study will involve no cost to you.

I have read and understood the information on this form. I have had the information on this form explained to me.

Signature of the interviewer:
Signature of the interviewee:

- **Block 1: Question about procedure & Teamwork**

1. There are different kinds of Interventional radiology, such as Angiography, Interventional Oncology, and Neuroradiology. What is the difference in the procedure compared to the other interventional therapies? Could you please describe the general procedure of angiography?
   Follow-up questions: Form the literature we know that the procedure could be general described: Create an angiographic run: Create reference image: Using reference image, Loop. (Check the answer with question 1, if there is a difference, ask about the differences)

Follow-up questions 2: Do the doctors always identify disease one by one? How they can guarantee not missing diseases?

2. How many doctors are involved in one operation? What are their roles and how they do the teamwork?
   Follow-up questions: In literature it was described as, 2 radiologist, 1 scrub nurse, a number of non-sterile nurse, one/two radiographer If there is a difference, ask why? Follow-up questions 2: How long does a surgery last?

3. Which console are manipulated by radiologists, and which was done by technologist? How they take their own responsibility? and How they communicate with each other? (Encourage them to describe a detailed example and the command name)
   Follow-up questions: Example in literature, "North, up down a bit more, open side a little bit for me, set up peri run? bring the table vertical up, get the table a bit higher, hang on, just to back up, the left side is the one we more interested in, mustn't come across. Get the legs a bit closer."
   How long does the communication take between Radiologist and Assistants? Can assistant understand the radiologists immediately? If not, what do they do to solve the problem?

4. Could you please describe the problem that you encountered before with as much details as you can? And how frequent?
   Follow-up questions: We found some problems mentioned by other research, have encounter any of them? If so, could you describe the details? And how frequent?

- Radiologists sometimes have to leave the operating room because they need to manipulate the non-sterile mouse in control room. There is a problem with keeping hands sterile for running the operation.
- Radiologist found their hands are occupied by holding the catheter and wires, but they want to adjust the console.
- Radiologists found their communication with assistants about precise adjustment is not successful after continuous repeating.
- Radiologists found themselves are very overloaded, as they need to take care of various agents at the same time, such as patient, managing the operation group, the devices, and etc.

5. (If radiologists mentioned there is a misunderstanding with assistant, ask this question) what kind of misunderstanding occurs during the progress? How do the radiologists solve the problem when misunderstanding happens?
   Encourage them to describe a detailed example and the professional terms they use.

6. Is the patient always conscious during the IR? do the radiologist communicate with patient? If so, what are they talking about?
   Encourage them to describe a detailed example and the professional terms they use.

7. What is your feeling after during and after doing an operation?

- **Block 2: Question about specific interactions**

1. What type of C-arm are you using in your practice?

2. (Asking the radiologists) Here are two control consoles from Siemens Artis Family (It might be different as the one you use in practice, it's ok, you can use them as references or describe the console you are using. Could you please describe me as detailly as possible how do you use the table console?

   (Asking the design experts) Here are two control consoles from Siemens Artis Family Could you please describe the designed function of each button and joysticks? Follow-up questions: What's the difference between these two consoles? Why you change it?

3. (Asking the radiologists) What kind of errors did the radiologists encounter with this console?

   (Asking the design experts) What kind of error prevention did you design?

4. What are the essential sub-screens shown on the Monitor? What else are additional sub-screen?
   Follow-up question: How does the radiology request or adjust or edit the display of the views? Could you please describe one example as detailed as possible of how you changing the views?

Encourage them describe how they do the specific interaction? Or how do technologists deliver reference image to the screen under their instructions? Or applying masks, make lines, and etc.

5. How does the radiography do the media injection? Is it required all the procedure? How was it started and stopped? Who is in charge of that?

6. When radiologist taking reference image, what parameters will be adjust to acquire request image? Such as how they can choose the scope of view and focus of view? What's the detail interaction? How they choose the reference image?

- **Block 3: Asking their subjective opinion about VI in specific tasks.**

We summarized some specific interaction during the angiography. As we are not experts, the name of interactions might not be exact as what you call them in practice, please feel free to tell the right name or asking me questions if you have any confusions.

In this section, you will give you opinion of voice interactions in each specific task, which of them do you think are proper to interaction with voice? Could you please speak out your reasons of why you pick this answer? What's else interaction are missing?

Interactions	Voice Command	Opportunity of voice interaction
Preparation		
Turn on the contrast injector		Very Possible 5 4 3 2 1 NotPossibleAtAll
Pre-set the display of the monitor	Such as, fluoroscopy in screen 1, ECG in screen 2	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Patient Information check		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Online conference connection		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Create angiographic run		
Adjust the position of C-arm	Arm1 left rotate 15 degree	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Adjust the table	up/left/right/down 1 cm	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Distribute views in monitor	Screen 1 load reference image	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Input contrast parameters	Dose, start time..	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Start/ Stop contrast injector in between		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Switch on/off different modes	Light on/ off; show lines, show notes	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Take reference image	Pedal on	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Change the scope of intensifier (zoom in/out	Fluoroscopy zoom in/out	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Lock/ Unlock the voice interaction		VeryPossible 5 4 3 2 1 NotPossibleAtAll
		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Making reference image: not beside computer rather directly beside table		
View reference image	go.../stop	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Choose reference image		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Delete reference image	Delete this one	VeryPossible 5 4 3 2 1 NotPossibleAtAll
Upload target reference image	Upload this one	VeryPossible 5 4 3 2 1 NotPossibleAtAll

Make 3-D model of the parts		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Addling mask		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Draw lines in the image		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Using reference image		
Zoom in/out reference image		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Adjust view (including 3D)		VeryPossible 5 4 3 2 1 NotPossibleAtAll
		VeryPossible 5 4 3 2 1 NotPossibleAtAll
Discover another disease		
Automatic discovering next disease		VeryPossible 5 4 3 2 1 NotPossibleAtAll
		VeryPossible 5 4 3 2 1 NotPossibleAtAll

2 APPENDIX INTERVIEW RESULTS

I. Interviewee I

Doctor J,
15 years of experience
expert in CVIT
C-arm type: GE single arm

1. Sub-classification of IR techniques in hospital

Interventional radiology can be divided as cardiac intervention (treating cardiovascular disease), Neuro-intervention (vascular disease in brain, spine and other neuro-vascular system), General Intervention (disease in the other parts including legs, organs, tumor problems).

2. Cardiovascular intervention procedure

a) First identify the disease occurs in right coronary artery (RCA) or right coronary artery (LCA) and the patient will swallow sedatives to calm down.
b) Then patient come to the hybrid operating room, the nurse help with disinfection.
c) The radiologists decide where to insert needles, it could be either entered from groin artery or brain artery when doing brain intervention.
d) Then insert wires to left artery. When the catheter arrives, (we could see it under fluoroscopy), we will start to inject contrast media. We call it angiography, and we will interpret the images to decide which part of the patient's artery are narrowed. Sometimes the anatomy of a particular patient might be very different. For example, there are congenital anomalies. Then we decide should we place stents or make vessel dilatation.
e) Angiography are triggered by foot pedal; contrast media will be injected at the same time. The created image will be show on the display and saved automatically. This manipulation could be either done inside the examination room, or by technologists sitting in the control room.
f) The injection of contrast media in cardiovascular are conducted **manually.** But in brain or tumor IR, we will **use pump.** The speed and flow rate can be pre-set by technologists.

3. Inspecting disease without missing or mistaken diagnosis

Normally cardiovascular CT scan will be conducted before IR, or dual source CT (DSCT). These images can be used to inspect disease in advance. Sometimes we will also directly conduct angiography to inspect disease. Some disease can also be found during the physical examination.

4. Team composition

The team of interventional group are normally consisting of 2 radiologists, 1 nurse, 1 cardiologist,1 technologist. The main radiologist is responsible for inserting catheters and placing stents. Another radiologist, who is usually younger or less experienced, will help with adjust the position, move tables (forward or backwards). Sometimes they will also inject medicine, or inject contrast media. The technologist in control room, are in charge of controlling and manipulating devices. He/she will input patients' name. gender, choose parameters, make the image clearer, save images, show reference image for radiologists.

5. Giving instructions to assistants.

There is some standard position that all radiologic talents know, so that sometimes assistants will adjust position without communication with radiologists. Both radiologists and assistants have the knowledge of the structure of cardiovascular. The table can also be adjusted manually forwards, backwards, left, right, up and down.

6. Problems in radiologists' point of view

- The problem that the radiologists do want to manipulate the devices on their own, but their hands are occupied by holding wires and catheters. In around 4/10 to 5/10 IR treatment I will encounter this problem. Because some patients' anatomies are different, so that the standard position are not suitable for them. When we encounter this problem, the radiologists will communicate with the assistant, and I re-position the C-arm.

- The problem that the radiologists will move to the non-sterile control room do happen a lot in practice. Because sometimes radiologists need to communicate with patients' family to ask for content of the next step.

- Another problem is that we are very tired, extremely tired after finishing an IR. The main reason relies on the lead aprons and lead hats that we should wear to protect us from radiation. It weights around 20 to 30 Jin (equals 10-15kg), and cover the body tightly. I usually stand to perform IR.

7. The functions that frequently used in table console

- moving tables
- adjust angles of C-arm
- zoom in / zoom out aperture.
- save reference images.
- go to standard position

The touchscreen is not frequently used. Because the technologists can do that faster and more precisely.

8. Displayed sub-screens

- Patient condition monitors: ECG/ blood pressure/ heart rate/ pulse / blood pressure

- Fluoroscopy
- Angiography images.
- Reference images such as previous CT scan, previous angiography images

9. Attitude to VI in IR

a) Regard as reasonable:
    - Adjust the position of C-arm: 5
    - Adjust the table:5
    - Arrange display: 5
    - Check patient info: 5
    - Add mask (the pre-set masks): 5
    - Adjust the dose: 5

b) Regard as impossible:
    - Trigger angiography images: 1

because the timing is very important. voice interaction will be definitely slower than foot pedal. but we do want to see it immediately, delay several seconds will be a huge problem.

II. Interviewee II

Doctor H,
Expert in Neuro-intervention
C-arm type: GE single arm argo730

0. The classification of IR

The intervention can be divided as neuro-intervention, general intervention which is normally tumor intervention, and cardio-intervention.

1. The procedure of neuro-intervention

This procedure is more complex and difficult than other type of intervention.
The problem(disease) can be pre-identified by CT scan normally. But sometimes the diagnose might be wrong. For example, we didn't find any disease during the intervention. But it happens rarely.
The procedure starts with Digital Subtraction Angiography (DSA), until the whole brain is inspected, we will interpret the images and decide the therapy. When we regard it is suitable to perform IR treatment, such as stent, we will ask the content of the patients' family. Then we will do angiography again, to reach the disease and treat it. The angiography takes around 30minutes, A treatment takes 2-3 hours normally.

2. How to give instructions

The technologist is sitting in the control room, we would give instructions or requirements to them. If they don't understand or do wrong manipulation, we will repeat the instructions. There is a control console of the C-arm and table in the examination room as well.

3. Team composition and responsibilities

The core team is consisting of radiologists, technologists, anesthetists …The anesthetists mostly perform local anesthesia of the patient (such as angiography, extracranial stents) but for disease like intracranial aneurysm, they will perform general anesthesia. Around half of the neuro-intervention will apply local anesthesia. Technologist are repressible for using computers in the control room, set-up and maintain the C-arm, call in needed images to display in examination room, prepare contrast pump.

4. Problems

- Radiologists will sometimes go to the control room directly to manipulate devices when technologist cannot place devices as needed. it also depends on how many radiologists attend the procedure. If there is only one radiologist, this problem happens more often, as least once per treatment. The problem of leaving sterile environment is not that severe. The sterile requirements are not so strict in the IR as open surgery. There are sterile covers to protect radiologist, so the problem is not that terrible.

- The problem of hand-occupied problem occurs quite often in IR. However, there are some solutions right now. For example, we use foot pedal to capture angiography images. And we will ask assistant to help with some other tasks. If the assistant didn't interpret my requirements correctly, I'll repeat the instructions. But assistant is trained in advance, in the most cases, only new assistant will encounter this problem.

- We are very tried when performing the IR procedure. We should wear over 10kg lead clothes for couple of hours. That's physical load, when we encounter complicate procedure, or challenging cases, the mental load is heavy too.

- The high-speed of development of technology are also challenging for us. we need to learn new technologies or be trained to catch up new trends.

5. Communication with patients

We are mostly focusing on the procedure, so that we didn't communicate with the patient very often, only when we need the cooperation form the patients, such as holding breath, no moving, etc.

6. Display

- images triggered by foot pedal
- roadmap
- sometimes contrast monitors

7. Attitudes to VI in IR

a) Regard as reasonable

- Adjust the position of C-arm: 5
- Set-up display: 5
- Ask help from experts remotely: make skype phone calls: 5
- Arrange display: 5
- Input contrast parameters:5
- Viewing images:5
- Adjust table: 4
- Zoom in/out reference image: 3

b) Regard as impossible:
- trigger angiography images: 1

The timing of foot pedal is critical, and sometimes last for a long time depends on the real situation.

III. Interviewee III

Doctor U,
Several years' work as radiologists
Several years' work as product manager in imaging area.

1. Intervention Procedure

Prepare the entrance of the needle, where you want to introduce the shift into the patient, in the clean environment. Do the puncher of the vessel, have the first view whether the shift is correctly placed. this you can do with the fluoroscopy system. After fluoroscopy is done, you already know where it is.
Later on, apply the contrast media, you can either follow in steps. Or you make a short walk by the movement of the table on your own, following the contrast media. Follow the contrast media where the contrast media go. This is the seldom case. The main case is you apply the contrast media, you have already the spot where the contrast media is backed to be located or to see. You do a lot of positioning (different views: left artery oblige and right artery oblige and spider view and four-chamber view and plane arterial angiography view.)
You have to do a lot of positioning the system, so you have optimal placement of where the contrast media flows in. We also have to apply the shelter because the contrast media in the heart, it has high contrast, but the lung in the side does not have a high reflection, so it is really white and therefore you have to apply your shelter, then you have a better differentiation inside the coronary arteries.
That was out basic procedure.
Sometimes we have to inflate the balloon, and also put in a stent inside. That was the main proves that the main stent is optimally be flatted, that it fits correctly to the border inside the vessel.
We have also sometimes view loop sequences, that before the procedure was done that in order to review the former procedure was done. In loop sequences because the clue that oh it was stop here, ok, as a closure of the coronary, then put the guide-wire following the …, in order to reopen the vessel. …. that was loops, we have done reference images, because I want to show my patient that

here was the coronary artery stenosisbefore the intervention then after the intervention. Then they can see there is such a difference, that I am cured now. That was reference image, we have sometimes print it out.

2. personal experience

I have performed coronary angiography procedures, either artery embolism or artery hemorrhage. I preferred use mouse to view the reference image after the procedure.

3. How to cooperate with an assistant

Go to the heart to the artic just above the arctic then find appropriate place. And along the catheter was positioned in the coronary arteries I have used the fluoroscopy. Or an assistant moves the table that I was always in a good position to see there the top of the catheter is placed along the movement. Because you do not see where the catheter is inside the body, therefore you use the fluoroscopy then help you find the correct pathway then you can go into the artery would be a deviation but you have to come along the way.

4. Cooperate with patient

And in here it is sometimes difficult to pass the entrance of the artery near the shoulder part. For that sometimes the patient has to move a little bit of the arm that the enters a bit manner. ask to hold breathe or breath. or report the doctor they are painful.

5. Adjust the contrast media

Yes, they have sometimes adjusted the speed. It depends on the magnetic of the vessel, how big the vessel is and how fast you want to have an access though. If you would put too much contrast media in to the coronaries then you can create a damage of the heart. Therefore, you have to pay attention this would not be too fast. For the rest area is not that critical.

6. Adjust the frame-rate

And also, sometimes you adjust the rate of the frames, which are acquired, so heart beating is in rate 60 bites per minutes, but the patient is a bit higher because they are anxious, so the heart bit is 80, b/m, children's heart beat 100 - 140 b/m. Then you have increased the frame rate of the acquisition system. Regularly it is 6, 15, 12 frames per seconds but that point we increase the 30, 60 frames for seconds. If the system is capable for 60 frames. That you can do more often the images instead of only 6 or only 15 images per second. That is the frame rate, how you acquire images that is this kind of loop themes or movie themes.

7. Do the doctors always identify disease one by one?

It depends on the skills of the radiologist, and the ordered procedure. you should do the procedure following the pre agreement. you should be at a level of trained

angiography interventional radiologist, and normally only those are allowed to perform those procedures that trained longer time. it can happen, if it is difficult case, or I do the diagnostic then I call my college, chef physician to the complicate case, and take over, and one is the assistant, and one is doing the intervention.
Or so it as 2 steps, first diagnose then go home and come back to be cured.

8. How they can guarantee not missing diseases?

It could happen, for example in the more complex situation, there are three arteries, you are inside the vessel, so you need a roadmap that is a fusion of precious CT image and an overlay of current situation, then you have a pathway. the reference image is not roadmap. reference image is image that are old images or CT images created by previous modalities or in old case. it has a good anatomy. sometimes you create reference image before and after the treatment, and show to the patient.

9. Team roles

It can be only one role, just one interventional radiology or radiologists. it should be a trained radiologist in intervention. Or in more complex case, it can be one radiologist + assistant + nurse +MTRA (technologist). The nurses would be the assistant or the MTRA, and they can help with the table movement, inject contrast media, hand over catheters and so on.

10. Time

It depends, a simple angiography could last in 10-15 minutes, but if intervention is involved, it’s about 30 minutes or longer.

11. Cooperate with assistants

the assistant are physicians well educated, so they will follow the contrast media, do the standard position or more. it would happen in some case that the position is not optimal, then it would not be precise. you would say: go a bit more. come down plus 5, plus 10.
commands example: 5 milligram medicine now, hyper IV, give me the ... 5 french, move the table up, old position, shelter in, reduce the frame. the misunderstanding of the commands not very often (once per month.)

12. Technologist

They sit outside. they will help type in medicine parameters, start on the table, emergency call for help.

13. other problem

5%-10%, I will leave the sterile environment and go to the control room to do it on my own.
10-15% that I found my hands are occupied and I want to adjust the devices on my own.

The systems are not integrated, for example. the Ultrasound are not visible, and not displayed. or the intergradation of reference image. the nurse would also misunderstand the materials that I want.

14. Attitudes to VI in IR

a) Regard as reasonable
- Set-up display: 5
- Ask help from experts remotely:5
- Adjust the position of C-arm: 5, I need the feedbacks
- Make a focus: 5
- Lock/ unlock voice commands: 5
- Switch on/off different modes: 5
- Input contrast parameters: 4, feedback is important
- Start / Stop contrast injector: 4
- Zoom in/out image: 4
- Make 3-D model of the parts: 4

b) Regard as impossible:
Take reference image:1

Generally, he thinks the voice in IR should be consider cautiously, because it is a delicate manipulation, and complicate environment, especially when there is continuous interactions, like adjusting table and c-arm, feedback is important in case any accident happened.

IV. Interviewee IV

Designer A, senior designer of C-arm devices
She introducing the functions in table console and how it works

V. Interviewee V

Dr R Expert in designing C-arm devices.

a) Regard as reasonable
- Setup displays: 4
- Make a focus point:4
- Input contrast parameters:4
- Start/stop contrast injector: 5
- Lock/ unlock voice commands: 5
- Choose reference image: 5
- Zoom in/out image: 4
- Make a focus point:4
- View reference image: 3
- Adding mask: 4
- Upload target reference image:3
- Annotations in image: 5

b) Regard as impossible:
    - Adjust the position of C-arm: 2, It is hard to giving out commands and make it precisely positioned as the radiologists wanted.
    - Adjust table: 2
    - Re-arrange display: 2, hard to define voice command
    - Make 3D model: 2
    - Turn on the contrast injector:2
    - Draw lines in image: 1, hard to define voice command

Generally, he thinks the voice command should be short and clear, which can be easily and correctly processed by computers. Otherwise the interaction Amy encounter more problems with error prevention.

VI. Interviewee VI

Designer I, USER EXPERIENCE Designer with experience in VI in medical environment

a) Regard as reasonable
    - Adjust the position of C-arm: 4
    - Make a focus: 4
    - Lock/ unlock voice commands: 4
    - Re-arrange display: 3
    - Input contrast parameters: 4
    - Switch on/off different modes: 4
    - Start / Stop contrast injector: 4
    - Zoom in/out image: 5

b) Regard as impossible:

Overview of the FB to VI

	H	J	U	I	D	M.	SD.	Remaks
	Radiologist			Designer				
Turn on the contrast injector					2	2,00	0,00	Defected
Set-up displays	5		5		4	4,67	0,58	Positive
Ask help from experts remotely	5		5			5,00	0,00	Positive
Check patient info		5				5,00	0,00	Defected
Adjust the position of C-arm	5	5	5	4	2	4,20	1,30	Conflict
Make a focus point			5	4	4	4,33	0,58	Defected
Adjust the table	4	5			2	3,67	1,53	Conflict
Arrange displays	5	5		3	2	3,75	1,50	Conflict
Input contrast parameters	5		4	4	4	4,25	0,50	Positive
Start/ Stop contrast injector			4	4	5	4,33	0,58	Positive
Switch on/off different modes			5	4	5	4,67	0,58	Positive
Take reference image	1	1	1			1,00	0,00	Negative
Zoom in / out image			4	5	5	4,67	0,58	Positive
Lock/ Unlock the voice interaction			5	4	5	4,67	0,58	Positive
Adjust dose		5				5,00	0,00	Defected
View reference image	5				3	4,00	1,41	Positive
Choose reference image					5	5,00	0,00	Defected
Delete reference image								
Upload target reference image					3	3,00	0,00	Defected
Make 3-D model of the parts			4		2	3,00	1,41	Neutral
Addling mask		5			4	4,50	0,71	Positive
Draw lines in the image					1	1,00	0,00	Defected
Zoom in/out reference images	3		4	5	4	4,00	0,82	Positive
Adjust view (including 3 D)			4			4,00	0,00	Defected
Automatic discovering								
Annotations in image					5			Defected

Remarks:
We didn't use the FB with only one score.
We marked the results in three categories, positive, negative, neutral, and conflict for further analysis.
Task "Make a focus point" is defected, because it is an assumed task, we didn't find reference for this task.

3 APPENDIX TESTING PROTOCOL

Consent form

Purpose of the study: The goal of this study is to explore the working performance of a voice interaction system in the image guiding procedure when different complex cases occurs.

Procedures: you will be asked to take part in 4 tasks and play the “spiral game” in different conditions with the guidance of the live stream camera, which means you cannot see the metal route and hand shank directly.

Recorded data: Our goal is to compare the performance (accuracy and effectiveness) in these three cases, therefore we will record your times of restart; how far you go; your real-time heart rate and your feedback via a questionnaire. We will also record the testing procedure in video or picture format for meta-analysis.

The goal of the testing is not to evaluate your skill or capacity of playing games but only for research purpose. We will also ask you some personal information to eliminate interference, they are:

Your gender: Male/ Female **Your age:**

Which hand is your dominate hand? Right/ Left

Alternative to Participation: Participation in this study is voluntary. You are free to withdraw or discontinue participation at any time.
Location: DesignLab
Confidentiality: Your personal data will be anonymized so that it cannot be trace back to you. Statistic results and during-testing pictures and videos will be used in the final research and final presentation. Your personal info won’t be revealed.
If you agree to join this study, please sign your name below.
-------I have read and understood the information on this form.
-------I have had the information on this form explained to me,

Subject’s Signature: Date:

Witness to Consent Procedures: Date:

Principal Investigator: Date:

Observation form

Participant: **Time:** **Date:**

Test A		**Start time:**	
Testing item	Unit	Amount	Marks
Errors (restart)	times		
Use voice commands	times		
Distance	mm		

Test B		**Start time:**	
Testing item	Unit	Amount	Marks
Errors (restart)	times		
Use voice commands	times		
Error occurs when using voice	times		
Distance	mm		

Test C		**Start time:**	
Testing item	Unit	Amount	Marks
Errors (restart)	times		
Use voice commands	times		
Error occurs when using voice	times		
Error occurs when asking questions	times		
Distance	mm		

Test D		**Start time:**	
Testing item	Unit	Amount	Marks
Errors (restart)	times		
Use voice commands	times		
Errors occurs when using voice	times		
Distance	mm		

Recruiting poster

After testing questionnaire

Part I: ASQ

Test A

1. Overall, I am satisfied with the ease of completing this task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

2. Overall, I am satisfied with support information (Instructions, camera views, documents) when completing the task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

Test B

1. Overall, I am satisfied with the ease of completing this task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

2. Overall, I am satisfied with the way (voice command) to call for new views.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

3. Overall, I am satisfied with support information (Instructions, camera views, documents) when completing the task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

Test C

1. Overall, I am satisfied with the ease of completing this task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

2. Overall, I am satisfied with the way (voice command) to call for new views.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

3. Overall, I am satisfied with support information (Instructions, camera views, documents) when completing the task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

Test D

1. Overall, I am satisfied with the ease of completing this task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

2. Overall, I am satisfied with the way (voice command) to call for new views.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

3. Overall, I am satisfied with support information (Instructions, camera views, documents) when completing the task.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

Part II: Open Questions

I. What do you think is the main reason that make you trigger the alarm? (multiple choice)

- I get distracted when using voice to adjust view.
- The people surrounding me attracts my attention.
- After I give out voice commands, the view I needed didn't show to me in time.
- There are too much views, and I can not take all of them into account.
- The device is too sensitive to trigger alarm
- I get used to watch my hand, so I move my attention out of the screen.
- My arms are so tired for holding the hand shank for a long-time.
- I just want to finish the game as fast as possible, I don't care the alarm.
- Other reason:

2. I frequently witching my attention among different views.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

3. I only focus on one view.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

4. I know which view I needed during the game and give out right voice commands.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

5. I think the offered four views are sufficient for completing the tasks.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

6. I fully focused on the game and didn't notice the surroundings.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

7. I can easily answer the question from the host without interrupting the game.

STRONGLY AGREE 1 2 3 4 5 6 7 **STRONGLY DISAGREE**

8. Which task do you think is the most difficult?

9. Which view do you think is the most helpful?

4 APPENDIX TESTING RESULTS

Comparison of test C with test B

Compare performances between condition B and C

Compare Errors	B	C	Diff(C-B)
	5	7	2
	6	8	2
	7	7	0
	7	9	2
	4	5	1
	8	5	-3
	4	5	1
	5	2	-3
	7	10	3
	3	6	3
	7	5	-2
	5	2	-3
	5	5	0
M	5,62	5,85	0,23
SD.	1,50	2,38	0,87
VA.	2,26	5,64	3,38

Compare distance	B	C	Diff (C-B)
	9,2	8,1	-1,1
	12,8	9,4	-3,4
	19,3	18,8	-0,5
	10,4	9,2	-1,2
	9,2	21,5	12,3
	9,9	5,3	-4,6
	8,3	11,3	3
	9,2	19	9,8
	14,1	6	-8,1
	9,1	9,6	0,5
	7,1	7,8	0,7
	4,5	14,8	10,3
	5,9	3,2	-2,7
M	9,92	11,08	1,15
SD.	3,79	5,73	1,94
VA.	14,36	32,86	18,50

t-Test: Paired Two Sample for Means

	Variable b	*Variable c*
Mean	5,615385	5,846154
Variance	2,25641	5,641026
Observations	13	13
Pearson Correlation	0,379114	
Hypothesized Mean	0	
df	12	
t Stat	-0,365148	
P(T<=t) one-tail	0,360678	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,721355	
t Critical two-tail	2,178813	

t-Test: Paired Two Sample for Means

	Variable b	*Variable c*
Mean	9,923077	11,07692
Variance	14,36026	32,85692
Observations	13	13
Pearson Correlation	0,215287	
Hypothesized Mean [	0	
df	12	
t Stat	-0,67609	
P(T<=t) one-tail	0,255902	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,511804	
t Critical two-tail	2,178813	

Paired t-test in performance between test B and test C. There was no significant difference in errors between test B (M = 5.62, SD = 1.50) and test C (M = 5.85, SD = 2.38, t (12) = 0.365, p < .05). Similarly, there was no significant difference in distances between test B (M = 9.92, SD = 3.79) and test C (M = 11.08, SD = 5.73, t (12) = -0.676, p < .05.) We failed to reject $H0_{er.bc}$ and $H0_{dif.bc.}$.

Compare experience between condition B and C

Compare ease	B	C	Diff (C-B)
	7	6	-1
	7	7	0
	3	3	0
	7	7	0
	6	5	-1
	6	6	0
	5	3	-2
	7	6	-1
	6	3	-3
	6	7	1
	4	3	-1
	6	5	-1
	6	5	-1
M	5,85	5,08	-0,77
SD.	1,21	1,61	0,39
VA.	1,47	2,58	1,10

Compare satisfication of VI	B	C	Diff(C-B)
	2	3	1
	4	3	-1
	3	2	-1
	1	4	3
	4	4	0
	3	2	-1
	2	3	1
	3	1	-2
	2	2	0
	3	3	0
	2	2	0
	7	3	-4
	1	1	0
M	2,85	2,54	-0,31
SD.	1,57	0,97	-0,61
VA.	2,47	0,94	-1,54

Compare satisfication of Info	Study B	Study C	Diff (C-B)
	5	3	-2
	3	5	2
	3	3	0
	5	4	-1
	4	4	0
	2	3	1
	6	2	-4
	1	1	0
	3	3	0
	6	6	0
	2	2	0
	4	4	0
	4	1	-3
M	3,69	3,15	-0,54
SD.	1,55	1,46	-0,09
VA.	2,40	2,14	-0,26

t-Test: Paired Two Sample for Means

	Variable b	*Variable c*
Mean	5,8461538	5,0769231
Variance	1,474359	2,5769231
Observations	13	13
Pearson Correlati	0,7761311	
Hypothesized Me	0	
df	12	
t Stat	2,7386128	
P(T<=t) one-tail	0,0089887	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,0179775	
t Critical two-tail	2,1788128	

t-Test: Paired Two Sample for Means

	Variable b	*Variable c*
Mean	2,8461538	2,5384615
Variance	2,474359	0,9358974
Observations	13	13
Pearson Correlation	0,2232574	
Hypothesized Mean Differenc	0	
df	12	
t Stat	0,6713451	
P(T<=t) one-tail	0,2573583	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,5147166	
t Critical two-tail	2,1788128	

t-Test: Paired Two Sample for Means

	Variable b	*Variable c*
Mean	3,6923077	3,1538462
Variance	2,3974359	2,1410256
Observations	13	13
Pearson Correlation	0,4272361	
Hypothesized Mean Difference	0	
df	12	
t Stat	1,2034433	
P(T<=t) one-tail	0,1260046	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,2520092	
t Critical two-tail	2,1788128	

Paired t-test in experience between test B and test C. There was a statistically significant difference in ease of task completion between test B (M = 5.85, SD = 1.21) and test C (M = 5.08, SD = 1.61, t(12) = -2,738, p < 0.05). Conversely, there was no significant difference in satisfaction regarding voice interaction between test B ($M = 2.85$, $SD = 1.57$) and test C ($M = 2.54$, $SD = 0.97$, $t(12) = 0.671$, $p < .05$). Likewise, there was no significant difference in satisfaction regarding support information between test B ($M = 3.69$, $SD = 1.55$) and test C ($M = 3.15$, $SD = 1.46$, $t(12) = 1.203$, $p < .05$). We can reject $H0_{ea.bc}$; however, we failed to reject $H0_{sa.v.bc}$ and $H0_{sa.i.bc.}$

Comparison of test D with test B

Compare performances between condition B and D

Compare errors	B	D	Diff (D-B)
	5	4	-1
	6	4	-2
	7	5	-2
	7	5	-2
	4	4	0
	8	8	0
	4	2	-2
	5	4	-1
	7	6	-1
	3	4	1
	7	4	-3
	5	6	1
	5	4	-1
M.	5,62	4,62	-1,00
SD.	1,50	1,45	-0,06
VA.	2,26	2,09	-0,17

Compare distance	B	D	Diff(D-B)
	9,2	5,8	-3,4
	12,8	19	6,2
	19,3	15,8	-3,5
	10,4	11	0,6
	9,2	17,9	8,7
	9,9	7,6	-2,3
	8,3	10,3	2
	9,2	17,8	8,6
	14,1	13,5	-0,6
	9,1	5,3	-3,8
	7,1	11,2	4,1
	4,5	6,1	1,6
	5,9	2,3	-3,6
M.	9,92	11,05	1,12
SD.	3,79	5,47	1,68
VA.	14,36	29,87	15,51

t-Test: Paired Two Sample for Means

	Variable b	*Variable d*
Mean	5,615385	4,615385
Variance	2,25641	2,089744
Observations	13	13
Pearson Correlation	0,655349	
Hypothesized Mean	0	
df	12	
t Stat	2,94392	
P(T<=t) one-tail	0,006141	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,012282	
t Critical two-tail	2,178813	

t-Test: Paired Two Sample for Means

	Variable b	*Variable d*
Mean	9,923077	11,04615
Variance	14,36026	29,86936
Observations	13	13
Pearson Correlation	0,560243	
Hypothesized Mean I	0	
df	12	
t Stat	-0,883136	
P(T<=t) one-tail	0,197255	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,394511	
t Critical two-tail	2,178813	

Paired t-test in performance between test B and test D. There was a statistically significant difference in error between test B ($M = 5.62$, $SD = 1.50$) and test D ($M = 4.62$, $SD = 1.45$, $t(12) = 2.944$, $p < .05$). Additionally, there was no significant difference in distances between test B ($M = 9.92$, $SD = 3.79$) and test D ($M = 11.05$, $SD = 5.47$, $t(12) = -0.883$, $p < .05$). We rejected H0er.bd, but we failed to reject H0dif.bd.

Compare experience between condition B and D

Compare ease	B	D	Diff (D-B)
	7	7	0
	7	6	-1
	3	4	1
	7	6	-1
	6	6,50	0,5
	6	5	-1
	5	5	0
	7	6	-1
	6	4	-2
	6	7	1
	4	4	0
	6	2	-4
	6	6	0
M.	5,85	5,27	-0,58
SD.	1,21	1,45	0,24
VA.	1,47	2,11	0,63

Compare satisfication of VI	B	D	Diff(D-B)
	2	4	2
	4	3	-1
	3	2	-1
	1	2	1
	4	4	0
	3	2	-1
	2	3	1
	3	1	-2
	2	2	0
	3	3	0
	2	2	0
	7	3	-4
	1	1	0
M.	2,85	2,46	-0,38
SD.	1,57	0,97	-0,61
VA.	2,47	0,94	-1,54

Compare satisfication of Info	B	D	Diff (D-B)
	5	3	-2
	3	5	2
	3	3	0
	5	4	-1
	4	4	0
	2	2	0
	6	3	-3
	1	1	0
	3	4	1
	6	6	0
	2	1	-1
	4	4	0
	4	3	-1
M.	3,69	3,31	-0,38
SD.	1,55	1,44	-0,11
VA.	2,40	2,06	-0,33

t-Test: Paired Two Sample for Means

	Variable b	*Variable d*
Mean	5,846154	5,269231
Variance	1,474359	2,108974
Observations	13	13
Pearson Correlat	0,498034	
Hypothesized Me	0	
df	12	
t Stat	1,538968	
P(T<=t) one-tail	0,074877	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,149755	
t Critical two-tail	2,178813	

t-Test: Paired Two Sample for Means

	Variable b	*Variable d*
Mean	2,846154	2,461538
Variance	2,474359	0,935897
Observations	13	13
Pearson Correlation	0,379116	
Hypothesized Mean Differenc	0	
df	12	
t Stat	0,923186	
P(T<=t) one-tail	0,18705	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,3741	
t Critical two-tail	2,178813	

t-Test: Paired Two Sample for Means

	Variable b	*Variable d*
Mean	3,692308	3,307692
Variance	2,397436	2,064103
Observations	13	13
Pearson Correlation	0,645482	
Hypothesized Mean Difference	0	
df	12	
t Stat	1,099853	
P(T<=t) one-tail	0,146486	
t Critical one-tail	1,782288	
P(T<=t) two-tail	0,292971	
t Critical two-tail	2,178813	

Paired t-test in experience between test B and test D. There was no significant difference in ease of task completion between test B ($M = 5.85$, $SD = 1.21$) and test D (M = 5.27, SD = 1.45, t (12) = 1.539, $p < 0.05$). There was also no significant difference in satisfaction regarding voice interaction between test B (M = 2.85, SD = 1.57) and test C (M = 2.46, SD = 0.97, t (12) = 0.923, $p < .05$). Similarly, there was no significant difference in satisfaction regarding support information between test B (M = 3.69, SD = 1.55) and test D (M = 3.31, SD = 1.44, t (12) = 1.100, $p < .05$). We can reject H0ea.bc, but we failed to reject $H0_{ea.bd}$, $H0_{sa.v.bd}$, and $H0_{sa.i.bd}$.

Comparison of test A with test B

Compare performances between condition A and B

Compare error	A	B	Diff (B-A)
	3	5	2
	6	6	0
	6	7	1
	7	7	0
	2	4	2
	2	8	6
	2	4	2
	5	5	0
	4	7	3
	2	3	1
	7	7	0
	0	5	5
	8	5	-3
M.	4,15	5,62	1,46
SD.	2,51	1,50	-1,01
VA.	6,31	2,26	-4,05

Compare distance	A	B	Diff (B-A)
	64,1	9,2	-54,9
	45,1	12,8	-32,3
	224,8	19,3	-205,5
	68,1	10,4	-57,7
	173,3	9,2	-164,1
	260,5	9,9	-250,6
	147,5	8,3	-139,2
	42,5	9,2	-33,3
	343,6	14,1	-329,5
	155,3	9,1	-146,2
	97,9	7,1	-90,8
	267,5	4,5	-263
	61,1	5,9	-55,2
M.	150,10	9,92	-140,18
SD.	98,71	3,79	-94,92
VA.	9744,49	14,36	-9730,13

t-Test: Paired Two Sample for Means

	Variable a	*Variable b*
Mean	4,1538462	5,6153846
Variance	6,3076923	2,2564103
Observations	13	13
Pearson Correlation	0,414592	
Hypothesized Mean I	0	
df	12	
t Stat	-2,260197	
P(T<=t) one-tail	0,0215978	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,0431955	
t Critical two-tail	2,1788128	

t-Test: Paired Two Sample for Means

	Variable a	*Variable b*
Mean	150,1	9,9230769
Variance	9744,4917	14,360256
Observations	13	13
Pearson Correlation	0,2577145	
Hypothesized Mean D	0	
df	12	
t Stat	5,1675187	
P(T<=t) one-tail	0,0001171	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,0002341	
t Critical two-tail	2,1788128	

Paired t-test results regarding error and distance between test A and test B. Results reveal significant difference in error and distance. $H0_{er.ab}$ and $H0_{dis.ab}$ were, therefore, rejected.

Compare experience between condition A and B

Compare ease	A	B	Diff (B-A)
	3	7	4
	5	7	2
	2	3	1
	5	7	2
	3	6	3
	2	6	4
	1	5	4
	5	7	2
	2	6	4
	3	6	3
	2	4	2
	1	6	5
	4	6	2
M.	2,92	5,85	2,92
SD.	1,44	1,21	-0,23
VA.	2,08	1,47	-0,60

Compare satisfication of Info	A	B	Diff (B-A)
	1	5	4
	2	3	1
	1	3	2
	1	5	4
	3	4	1
	1	2	1
	1	6	5
	1	1	0
	1	3	2
	4	6	2
	4	2	-2
	4	4	0
	1	4	3
M.	1,92	3,69	1,77
SD.	1,32	1,55	0,23
VA.	1,74	2,40	0,65

t-Test: Paired Two Sample for Means

	Variable a	*Variable b*
Mean	2,9230769	5,8461538
Variance	2,0769231	1,474359
Observations	13	13
Pearson Correlati	0,6117586	
Hypothesized Me	0	
df	12	
t Stat	-8,874888	
P(T<=t) one-tail	6,398E-07	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	1,28E-06	
t Critical two-tail	2,1788128	

t-Test: Paired Two Sample for Means

	Variable a	*Variable b*
Mean	1,9230769	3,6923077
Variance	1,7435897	2,3974359
Observations	13	13
Pearson Correlation	0,1097357	
Hypothesized Mean Difference	0	
df	12	
t Stat	-3,319764	
P(T<=t) one-tail	0,0030564	
t Critical one-tail	1,7822876	
P(T<=t) two-tail	0,0061129	
t Critical two-tail	2,1788128	

Paired t-test results of ease of task completion, and satisfaction regarding support information between test A and test B. Results reveal significant differences.

An overview of the times of using different voice commands, and the means, the sum of total times in each test.

	Give out voice commands														
	Test B					Test C					Test D				
	F	M	L	T	Sum	F	M	L	T	Sum	F	M	L	T	Sum
P0	7	0	1	0	8	8	0	0	0	8	9	0	1	0	10
P1	9	0	0	0	9	7	0	0	0	7	12	0	0	0	12
P2	10	0	0	0	10	7	0	0	0	7	9	0	0	0	9
P3	8	0	1	1	10	7	0	0	0	7	9	0	1	1	11
P4	4	0	2	2	8	4	0	2	1	7	4	0	2	0	6
P5	9	0	2	1	12	4	0	1	1	6	5	0	1	1	7
P6	7	0	0	0	7	4	1	1	1	7	10	0	0	0	10
P7	7	0	2	0	9	7	0	1	2	10	9	0	1	1	11
P8	9	0	0	0	9	8	0	0	0	8	11	0	0	0	11
P9	10	0	0	0	10	8	0	0	0	8	10	0	0	0	10
P10	5	0	1	1	7	5	0	3	2	10	6	0	1	1	8
P11	16	0	0	0	16	13	0	0	0	13	16	0	0	0	16
P12	4	0	2	1	7	4	0	1	2	7	5	0	3	2	10
M	8,08	0,00	0,85	0,46	**9,38**	6,62	0,08	0,69	0,69	**8,08**	8,85	0,00	0,77	0,46	**10,08**
SD	3,15	0,00	0,90	0,66	2,47	2,53	0,28	0,95	0,85	1,89	3,29	0,00	0,93	0,66	2,47

We also compared the usage of the real-time front view (F), the static front view as a routemap (M), the real-time left view (L), and the real-time top view (T) (see Table 5). Overall the voice commands are used most often in test D (M = 10.08, SD = 2.47), then in the test B (M = 9.38, SD = 2.47), followed by test C (M = 8.08, SD = 1.89). The front view is used most often in all three conditions, and the front view accounts for 86.1% usage in test B, 81.9% in test C, and 87.8% in test D.

Results of each participant

P0	**29-11-2017**	**Study A**		**Study B**				**Study C**				**Study D**				**Remarks**
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		11:06	11:08	11:23:32		11:25:32		11:30:59		11:32:59		11:36:18		11:38:18		
C	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Rely on front view
D				7	0	1	0	8	0	0	0	9	0	1	0	
	Gaze movement			8				13				26				Less attention switch
	Error when asking Q							2								SC/ most errors
	Error when using V			0				3				2				SC/ most errors
	Error	3		5				7				4				SC / most errors
	Distance	64.1		9.2				8.1				5.8				The worse effectiveness is SD
	Heart rate															
	Special conditions	P0 is left-hand dominant, so he paly the game from left to right. The flexibility is worse than right-hand dominant participants, because of the location of the hole. P0 looks inside the box to adjust his arm (Answer F) , which are not allowed in the testing. But he said it's just for going back to the start point, so that we still regard his record effective.														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P0	3	1	7	2	5	6	3	3	7	4	3					X	X				5	2	5	1	1	3	C	F	
Tip																													

P1	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		11:55	11:57	12:07:22		12:09:22		12:19:03		12:21:03		12:15:02		12:17:02		
D	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Rely on front view
C				9	0	0	0	7	0	0	0	12	0	0	0	
	Gaze movement			8				12				6				Less attention switch
	Error when asking Q							2								50% error in 4 Q
	Error when using V			0				0				1				
	Error	6		6				8				4				SC raises most errors
	Distance	45.1		12.8				9.4				19.0				SD is better because of learning effect
	Heart rate															
	Special conditions	1. Think Spiral game is difficult 2. Get distracted during the testing (answer A/B), less-frequent switching attention,														

After testing questionnaire																														
	Study A		Study B			Study C			Study D			Open question																		Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9		
P1	5	2	7	4	3	7	3	5	6	3	5	X	X		X			X			6	2	5	3	3	6	C	F		
Tip																														

P2	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		12:16	12:18	12:46:07		12:48:07		12:50:50		12:52:50		12:56:12		12:58:12		
C	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Rely on front view
D				10	0	0	0	7	0	0	0	9	0	0	0	
	Gaze movement			29				16				19				frequent attention switch
	Error when asking Q							0								no error in 4 Q
	Error when using V			2				0				0				Less error using V
	Error	6		7				7				5				
	Distance	224.8		19.3				18.8				15.8				SB is better because of learning effect
	Heart rate															
	Special conditions	3. P2 is good at playing this game, she is trying to keep all the views into account and switch attentions between views, but she still rely on the front view.														

After testing questionnaire																														
	Study A		Study B			Study C			Study D			Open question																		Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9		
P2	2	1	3	3	3	3	2	3	4	2	3	X			X	X		X			5	2	2	6	1	4	C	F		
Tip																														

P3	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		13:15	1:17	13:31:52		13:33:52		13:44:14		13:46:14		13:39:25		13:41:25		
D	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Rely on front view
C				8	0	1	1	7	0	0	0	9	0	1	1	
	Gaze movement			35				28				35				Frequent attention switch
	Error when asking Q							6								150% error in 4 Q
	Error when using V			3				0				1				Less error using V
	Error	7		7				9				5				
	Distance	68.1		10.4				9.2				11.0				SC worse
	Heart rate															
	Special conditions	4. P3 perform worse in the Study C, she also give the 7 in the Open Q section 5. Although the gaze movement show that she frequently switch attention between views, but she don't think so and give 2 to the Q2 in Open Q														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P3	5	1	7	1	5	7	4	4	6	2	4	X			X	X		X	X		6	2	2	5	2	7	C	F	
Tip	I: View is not very clear for me, seems that the front view makes great sense for operation																												

P4	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		16:26	16:28	16:38:33		16:40:33		16:42:46		16:44:46		16:50:31		16:52:31		
C	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Make use of all views
D				4	0	2	2	4	0	2	1	4	0	2	0	
	Gaze movement			12				12				12				normal attention switch
	Error when asking Q							0								0 error in 4 Q
	Error when using V			0				0				0				0 error using V
	Error	2		4				5				4				
	Distance	173.3		9.2				21.5				17.9				SB worse
	Heart rate															
	Special conditions	6. P4 perform very well , he regard himself frequent switch attention between views in the testing, both Study C and D didn't make a obvious difference in his performance. And because of learning effect, SB seems the most difficult one for him. (Q8)														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P4	3	3	6	4	4	5	4	4	6, 5	4	4										2	6	2, 5	1	1	1	B	F	
Tip	I: Sticker are slippery, top camera has a delay																												

P5	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		18:12	18:14	18:22:46		18:24:46		18:32:44		18:34:44		18:39		16:52:31		
C	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Make use of all views
D				9	0	2	1	4	0	1	1	5	0	1	1	
	Gaze movement			29				21				17				frequent attention switch
	Error when asking Q							1								1 error in 4 Q
	Error when using V			1				2				0				SC worse
	Error	2		8				5				8				
	Distance	260.5		9.9				5.3				7.6				SB worse
	Heart rate															
	Special conditions	7. P5 also perform quite good, she didn't trigger many alarms. She try to take all views into account. 8. Although she give the highest grade to the Study C, and her performance is worst in Sc, but she still regard D as most difficult one. The reason we proposed is that Sd do required more effort to control the task, but as in Sc, the interaction came from outside, so they couldn't control the condition, so the feel less burdenin Sc														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P5	2	1	6	3	2	6	2	3	5	2	2			X		X	X	X			3	5	3	3	1	6	D	F	
Tip	I: Game too complex, route is too curved																												

P6	29-11-2017	Study A		Study B				Study C				Study D				Remarks
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		16:53	16:55	18:58:56		19:00:56		19:09:54		19:11:54		19:03:16		19:05:16		
D	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Trying to make use of all views
C				7	0	0	0	4	1	1	1	10	0	0	0	
	Gaze movement			36				22				53				Frequent attention switch
	Error when asking Q							0								0 error in 4 Q
	Error when using V			2				2				0				0 error using V
	Error	2		4				5				2				
	Distance	147.5		8.3				11.3				10.3				SB worse
	Heart rate															
	Special conditions	9. P6 also reply on the front view mostly, but she is trying to take all the views into account, so when SD, she show high gaze movement. 10. We regard the error are more valuable to indicate performance, as each time when user trigger alarm, they should go back to the start, so the distance also depends on the how many time left since last error. 11. The Errors do not correspond to the results in subjective feedback. The error occurs most in SC, but P6 give the highest ease in Sc (the reason could be the learning effect), but when asking which task is the most difficult one, P6 also choose SC.														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P6	1	1	5	2	6	3	3	2	5	3	3			X	X			X			5	3	2	6	1	2	C	F	
Tip																													

P7	**30-11-2017**	**Study A**		**Study B**				**Study C**				**Study D**				**Remarks**
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		13:17	13:19	13:27:04		13:29:04		13:30:40		13:32:40		13:34:41		13:36:41		
C	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Trying to make use of
D				7	0	2	0	7	0	1	2	9	0	1	1	all views
	Gaze movement			43				37				50				Frequent attention switch
	Error when asking Q							1								1 error in 4 Q
	Error when using V			1				1				2				
	Error	5		5				2				4				
	Distance	42.5		9.2				19.0				17.8				SB worse
	Heart rate															
	Special conditions	12. There is sunshine, which barriers the participants seeing the views. 13. The left view overlapped very much in this testing, she want right view 14. She think the learning effect influence her performance.														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P7	5	1	7	3	1	6	1	1	6	1	1			X				X			3	7	2	6	1	7	D	F	
Tip	5:need right view																												

P8	**30-11-2017**	**Study A**		**Study B**				**Study C**				**Study D**				**Remarks**
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		14:08	14:10	14:18:33		14:20:33		14:28:18		14:30:18		14:24:09		14:26:09		
D	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Rely on front view
C				9	0	0	0	8	0	0	0	11	0	0	0	
	Gaze movement			29				38				24				Frequent attention switch
	Error when asking Q							4								4 error in 4 Q
	Error when using V			1				0				3				
	Error	4		7				10				6				
	Distance	343.6		14.1				6				13.5				Worse in Sc.
	Heart rate															
	Special conditions	15. Also because learning effect give most ease in Sc, but still regard Sc most difficult in Open Q, and worse performance in Sc														

After testing questionnaire																													
	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P8	2	1	6	2	3	3	2	3	4	2	4	X			X			X			6	2	3	4	3	4	C	F	
Tip	I: lost patience after several alarm																												

P9	30-11-2017	Study A		Study B				Study C				Study D				Remarks
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		15:10	15:12	15:20:55		15:22:55		15:24:47		15:26:47		15:29:50		15:31:50		
C	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Rely on front view
D				10	0	0	0	8	0	0	0	10	0	0	0	
	Gaze movement			25				27				21				Frequent attention switch
	Error when asking Q							2								2 error in 4 Q
	Error when using V			1				2				1				
	Error	2		3				6				4				
	Distance	155.3		9.1				9.6				5.3				
	Heart rate															
	Special conditions	16. P9 think more close-up view might help 17. P9 think rely on the front view, and she wants the front view always existing.														

After testing questionnaire																														
	Study A		Study B			Study C			Study D			Open question																		Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9		
P9	3	4	6	3	6	7	3	6	7	3	6				X						3	5	6	6	4	6	D	F		
Tip	I: Front view disappear too soon																													

P10	30-11-2017	Study A		Study B				Study C				Study D				Remarks
A	**Time**	**Start**	**End**	**Start**		**End**		**Start**		**End**		**Start**		**End**		
B		16:08	16:10	16:18:27		16:20:27		16:26:31		16:28:31		16:23:02		16:23:02		
D	**Voice command**			**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	**F**	**M**	**L**	**T**	Trying to make use of all views
C				5	0	1	1	5	0	3	2	6	0	1	1	
	Gaze movement			34				30				35				Frequent attention switch
	Error when asking Q							2								2 error in 4 Q
	Error when using V			2				1				0				
	Error	7		7				5				4				
	Distance	97.9		7.1				7.8				11.2				Sb worse
	Heart rate															
	Special conditions	18. P10 will stay in the safe area to avoid the missing of front view 19. P10 think there is a learning effect 20. Because of learning effect, P10 give Study C most ease, however in Open Q8 he regard Study C as difficult. The worse performance in SB														

After testing questionnaire																														
	Study A		Study B			Study C			Study D			Open question																		Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9		
P10	2	4	4	2	2	3	2	2	4	2	1				X			X			2	5	2	3	2	4	C	F		
Tip																														

P11	30-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		17:49	17:51	18:02:50		18:04:50		18:06:38		18:08:38		18:10:40		18:12:40		
C	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Rely on front view
D				16	0	0	0	13	0	0	0	16	0	0	0	
	Gaze movement			13				14				5				less attention switch
	Error when asking Q							1								1 error in 4 Q
	Error when using V			2				1				2				
	Error	0		5				2				6				
	Distance	267.5		4.5				14.8				6.1				Sb worse
	Heart rate															
	Special conditions	21. P11 is very tired when doing the testing, so he is not satisfied with voice, the reason could be the dealy(Open Q C) 22. P11 show less attention switching and reply on front view (Open Q 2,3) 23. P11 think wearing jacket makes movement difficult.														

After testing questionnaire

	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P11	1	4	6	7	4	5	3	4	2	3	4		X	X							7	1	1	1	3	3	B	F	
Tip	I: missing depth view																												

P12	30-11-2017	Study A		Study B				Study C				Study D				Remarks
A	Time	Start	End	Start		End		Start		End		Start		End		
B		18:38	18:40	18:53:50		18:55:50		19:04:58		19:06:58		18:58:52		19:00:52		
D	Voice command			F	M	L	T	F	M	L	T	F	M	L	T	Trying to make use of all views
C				4	0	2	1	4	0	1	2	5	0	3	2	
	Gaze movement			28				17				18				less attention switch
	Error when asking Q							0								slowly
	Error when using V			0				1				3				
	Error	8		5				5				4				
	Distance	61.1		5.9				3.2				2.3				
	Heart rate															
	Special conditions	24. P12 is very tired and hungry when doing the testing. He get distracted by the surroudings 25. P12 feel hot, because wearing the jacket														

After testing questionnaire

	Study A		Study B			Study C			Study D			Open question																	Conclusion
	1	2	1	2	3	1	2	3	1	2	3	A	B	C	D	E	F	G	H	I	2	3	4	5	6	7	8	9	
P12	4	1	6	1	4	5	1	1	6	1	3			X				X			5	3	2	5	6	7	C	F	
Tip	I: hungry; 5 show all the views at the same time.																												