HCI in interactive segmentation
Human-computer interaction in interactive segmentation of CT images for radiotherapy

Ramkumar, Anjana

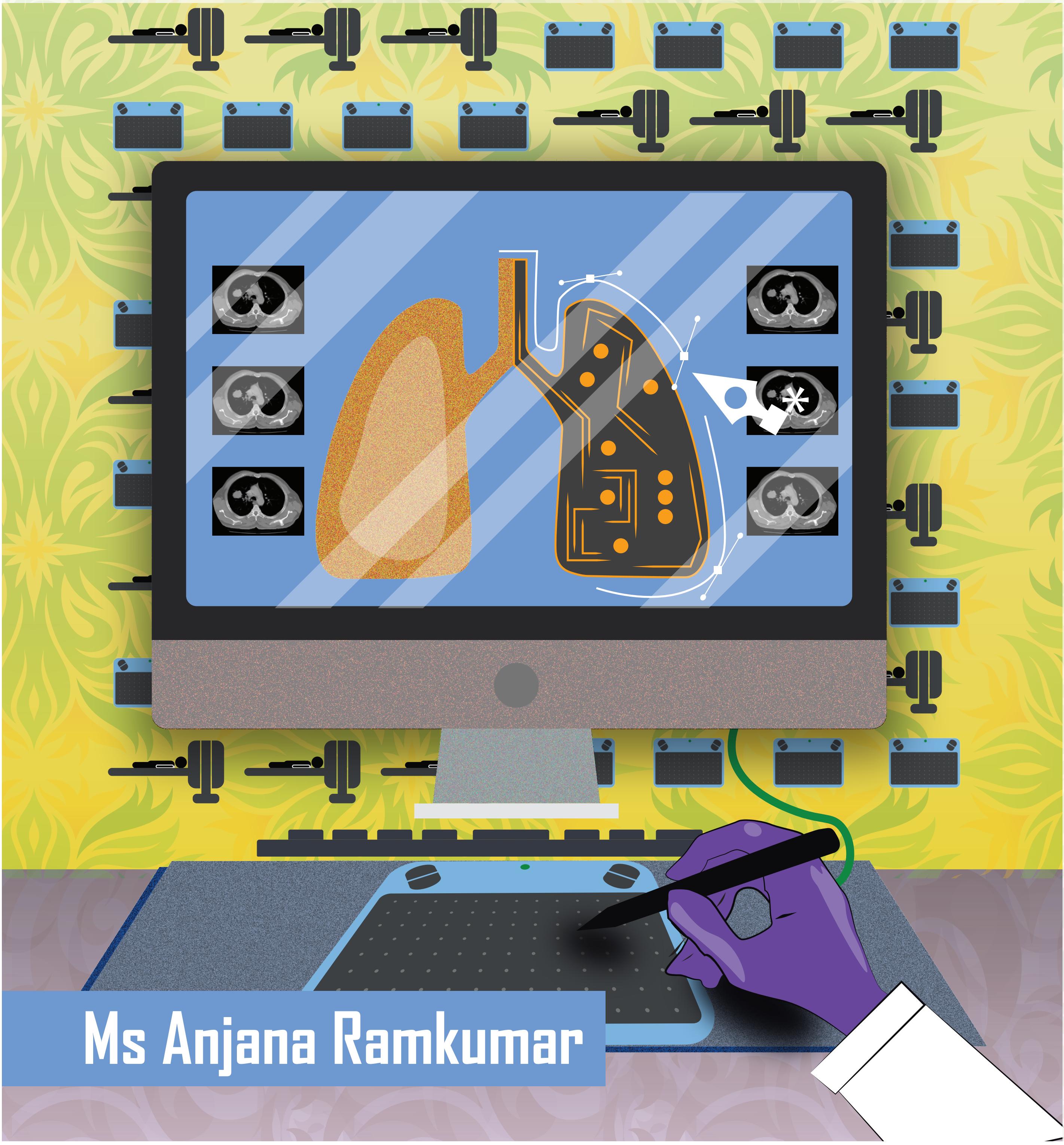
DOI
10.4233/uuid:0f0259f1-0c33-442f-b851-86a846e736fc

Publication date
2017

Document Version
Final published version

Citation (APA)
Ramkumar, A. (2017). *HCI in interactive segmentation: Human-computer interaction in interactive segmentation of CT images for radiotherapy*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0f0259f1-0c33-442f-b851-86a846e736fc

HCI IN INTERACTIVE SEGMENTATION



Ms Anjana Ramkumar

HCI IN INTERACTIVE SEGMENTATION

HUMAN-COMPUTER INTERACTION IN INTERACTIVE SEGMENTATION
OF CT IMAGES FOR RADIOTHERAPY

Proefschrift

Ter verkrijging van de graad van doctor Aan de Technische Universiteit Delft, Op gezag van de Rector Magnificus Prof. ir. K.C.A.M Luyben Voorzitter van het College voor Promoties, In het openbaar te verdedigen op maandag 10 april 2017 om 15:00 uur

door Anjana RAMKUMAR

Medical Physicist, University of Surrey, UK

Geboren te Renukoot,Uttar Pradesh, India

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr. W.J.Niessen
Prof.dr. P.J. Stappers
Copromotor:
Dr. Y. Song

Samenstelling promotiecommissie:
Rector Magnificus, Voorzitter
Prof. dr. W.J.Niessen Technische Universiteit Delft and Erasmus Medical Center, promotor
Prof. dr. P.J. Stappers Technische Universiteit Delft, promotor
Dr. Y. Song Technische Universiteit Delft, copromotor

Onafhankelijke leden:
Prof.dr.ir R.H.M Goossens Industrieel Ontwerpen, TU Delft
Prof.dr.ir. J.M.P Geraedts Industrieel Ontwerpen, TU Delft
Prof.dr.T.Brunner, Department of Radiation oncology, University Medical Centre Freiburg, Germany
Dr.S.D.Olabarriaga, Amsterdam Medical Centre, The Netherlands
Prof.dr.G.W.Kortuem Industrieel Ontwerpen, TU Delft, reservelid

Anjana Ramkumar

This research is funded by the European Commission (FP7-PEOPLE-2011-ITN) under grant agreement PITN-GA-2011-290148

Layout: Gopikrishna J and Dr.Vijay Harish S
Cover: Gopikrishna J and Dr.Vijay Harish S

ISBN: 978-94-92516-47-3

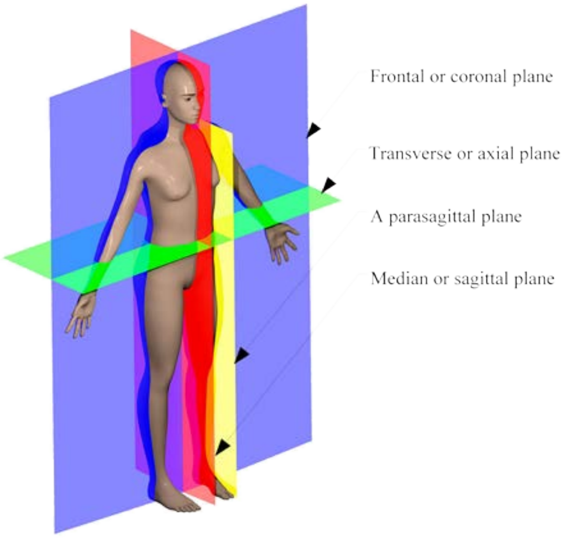
To my family

Table of Contents

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
BG	Background
CT	Computed Tomography
CTV	Clinical Target Volume
FG	Foreground
GOMS	Goals, Operators, Methods, and Selection rules
GTV	Gross Tumour Volume
HCI	Human-Computer Interaction
HF	Human Factors
ITV	Internal Target Volume
NASA-TLX	NASA Task Load Index
OAR	Organs At Risk
PTV	Planning Target Volume
RT	Radiotherapy
ROIs	Region Of Interests
RSI	Repetitive Strain Injury
UE	Usability Engineering
UI	User Interface
UID	User Input Device
TV	Target Volume

Anatomical planes

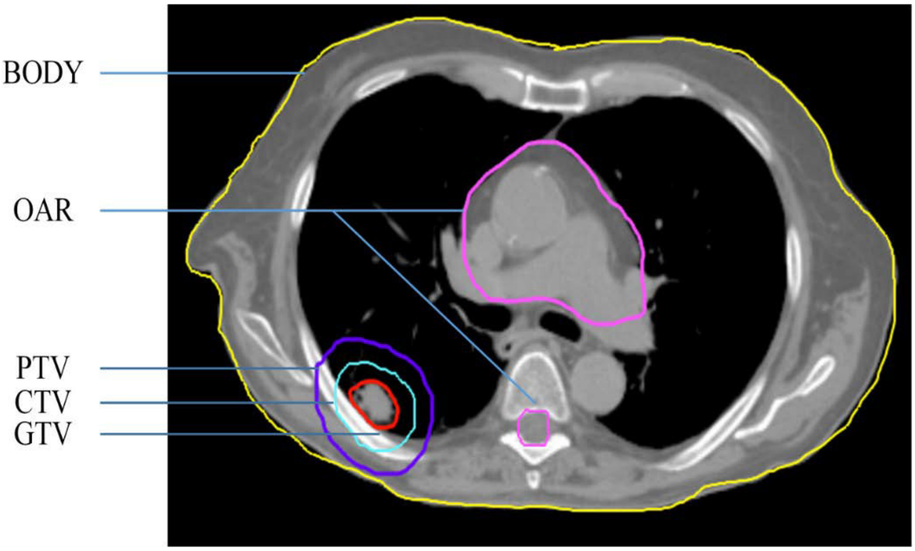


Anatomical planes

Figure adapted from [ANAL2017] under Creative Commons Attribution-ShareAlike License

Transverse or axial plane	The axial plane (also called the transverse plan, horizontal plane, or transaxial plane) is an imaginary plane that divides the body into superior and inferior parts
Median or sagittal plane, and parasagittal plane	The sagittal plane or median plane (longitudinal, anteroposterior) is a plane that divides the body into left and right.
Frontal or coronal plane	The coronal plane or frontal plane (vertical) divides the body into dorsal and ventral (back and front, or posterior and anterior) portions.

1 Introduction



This Chapter is based on the following publications:

Anet Aselmaa, Richard HM Goossens, Anne Laprie, Soléakhéna Ken, Tobias Fechter, **Anjana Ramkumar**, Adinda Freudenthal.
"Workflow analysis report". Retrieved from
http://summer-project.eu/wp-content/uploads/2013/10/D2.1_Workflow_analysis_report_WEB.pdf, 2013.

Anjana Ramkumar, Jose Dolz, Hortense A.Kirisli, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle, Laurent Massoptier, Edit Varga, Pieter Jan Stappers, Wiro J.Niessen and Yu Song.
"User Interaction in Semi-Automatic Segmentation of Organs at Risk: a Case Study in Radiotherapy".
Journal of Digital Imaging, 29(2), 264-277, 2016.

1.1 Radiotherapy

Cancer is one of the leading causes of death and in 2012, there were 8.2 million cancer deaths worldwide [WHO2017]. Radiotherapy (RT) is the treatment that involves the use of high energy radiations to destroy cancer cells in order to shrink tumours [NHS2016]. Its effectiveness is achieved by damaging the tumour cell's DNA so that these are unable to reproduce themselves. For instance, in a recent study, Corradini [CORR2015] indicated that in breast cancer management, 10-year overall survival rates were 55.2% with surgery alone vs. 82.2% when followed with postoperative radiotherapy ($p<0.001$). In the Netherlands, approximately 48% percent of cancer patients were treated by RT [SLOT2003, GRAU2014].

Two types of RT are used in clinical practice: external RT and internal RT, depending on if the radiation is given from outside or inside the human body. Though internal RT[UCLA2016] can deliver high level dose to the tumour in a more precise manner, it can only be applied to a few organs. Compared to internal RT, external RT can be applied to treat tumorous cells in/around nearly any organs. Figure 1.1 illustrates the setup of an external RT treatment. The patient undergoing the treatment is lying on the couch of the machine. A radiation therapist or a medical technologist, aligns the patient as per the treatment plan under the gantry of the treatment machine. Once the position is set, the planned treatment dose is delivered to the patient via the gantry head. The research presented in this thesis mainly focuses on the external radiotherapy, thus in the following, the abbreviation RT refers to external radiotherapy.

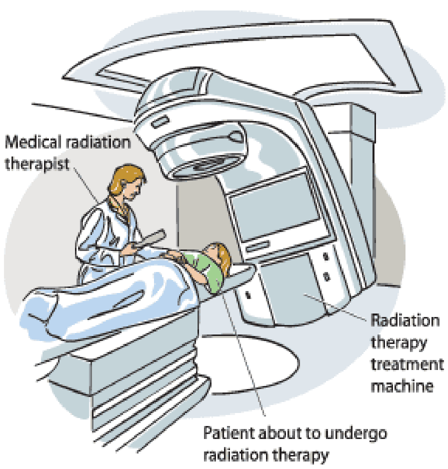


Figure 1.1: Setup of an external radiotherapy treatment (Courtesy of MedBroadcast [MEDB2016])

1.2 The workflow of the radiotherapy treatment

In clinical practice, RT has a lengthy workflow and it usually takes several days from planning to treatment [ASEL2013]. The general workflow of an external RT treatment is shown in Fig.1.2. Out of the four major steps, i.e., diagnosis, treatment planning, treatment and post-treatment follow-up, treatment planning is an important step as well as a complicated process which involves many stakeholders and tasks. The complexity lies in the fact that radiotherapy must be personalized for each patient.

Once the patient has been diagnosed and the treatment plan possibilities have been discussed in a multi-disciplinary meeting, and if RT has been suggested as (part of) the treatment plan, the patient comes to the RT consultancy. During consultancy, the process of RT and the steps involved in it are explained. The next step is to gather all needed data. For all cases a planning Computed Tomography (CT) scan is performed, and if required other materials such as immobilization system, gating training, etc. are prepared. Once all information about the patient and the tumour has been gathered, the planning of the treatment can start. If needed, images from different modalities are registered together for the delineation task. In the delineation, different target volumes with a margin around the tumour and the organs at risk (OAR) are contoured by oncologists on the available images, sometimes with the help of computational algorithms. The planned doses and limitations of doses for the tumour and the organs are then defined as well. The last step before the treatment is to create and validate a dose plan that is covering the tumour as prescribed and spares the OAR as much as possible. It is worth mentioning that in a tumour treatment plan, RT is often not the only method. Before, during or after the RT, other treatment methods such as chemotherapy, surgery, etc. may be applied. Those treatments may influence the general workflow of RT as well.

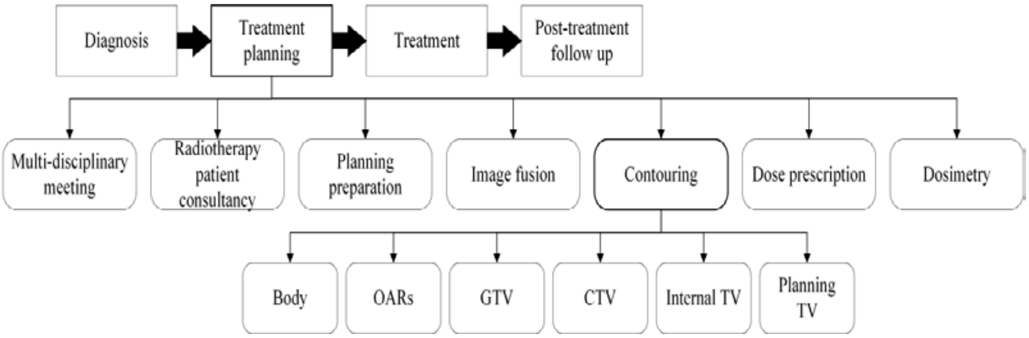


Figure 1.2: Radiotherapy workflow (GTV-Gross Tumour Volume, CTV-Clinical Target Volume, TV-Target Volume)

Based on the treatment planning, the treatment position is validated by an oncologist in order to ensure that there is no or limited deviation from the planned position. The medical physicists then verify the plan and makes necessary adjustments with the agreement of oncologist. During the treatment, a radiation therapist/medical technologist sets up the patient and the equipment, and treatment is delivered to the patient. There are weekly follow-up meetings during the treatment to evaluate the intermediate outcomes and/or complications. After the complete treatment, there will be post-treatment follow-up meetings to evaluate the successfulness of the treatment.

1.3 The contouring task

Contouring, also referred as delineation or segmentation, is an important step in the RT workflow where objects of interest are isolated from the background in order to make the representation of a volumetric image stack more meaningful and easier for analysis [OLAB2001]. The contouring step of the radiotherapy can be simplified into the following steps:

1. Delineating the body;
2. Delineating the OARs;
3. Delineating the gross tumour volume (GTV) which indicates the macroscopic tumour;
4. Delineating the clinical target volume (CTV) which indicates the microscopic tumour, i.e. infiltration;
5. Delineating the internal target volume (ITV) which indicates the expected movement of CTV during the treatment. ITV is drawn only for a few types

of cases where the tumour has a relatively large movement during the treatment, e.g., tumours in the lung;

6. Delineating the planned target volume (PTV) which is defined by setting margins to accommodate positioning errors in the treatment.

Figure 1.3 shows A CT image with multiple contours which represent the body, the OAR and GTV, CTV and PTV of the tumour. This thesis focuses on designing a more effective and efficient human-computer interactions (HCI) for delineating OAR for radiotherapy. However, many of the results can be applied to tumour contouring and other contouring tasks as well.

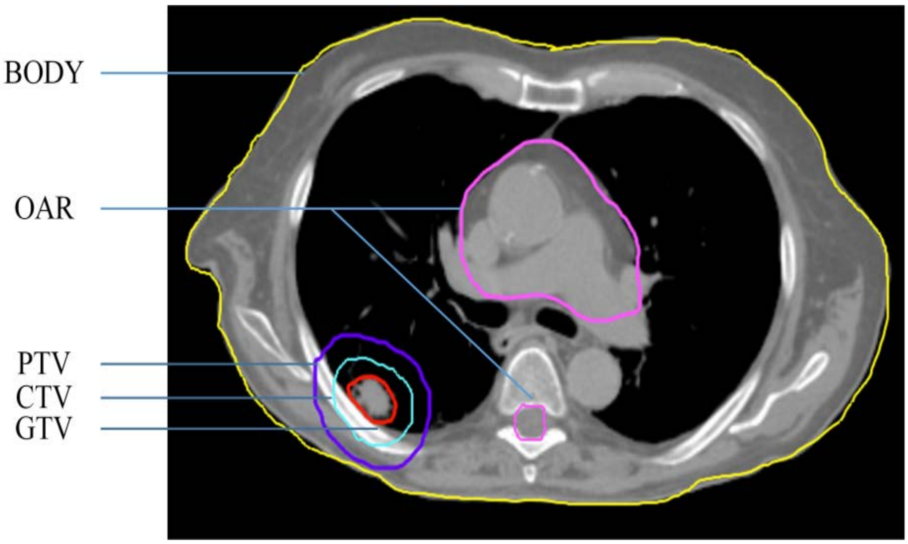


Figure 1.3: A CT image with multiple contours which represent the body, the OAR and GTV, CTV and PTV of the tumour

The contouring task is the weakest link in the search for accuracy in radiotherapy [NJEH2008]. Errors (From human and machine) introduced in the contouring task lead to systematic errors which cannot be eliminated in the remainder of the steps.

Generally, there are three different ways of performing medical image segmentation tasks: automatic, semi-automatic and manual segmentation methods. Among those methods, fully automated, operator independent segmentation methods have limited applications due to the inhomogeneity of anatomical structures, and low contrast and noise in medical images. Sims et al. [SIMS2009] concluded that careful review and manual editing were always required for most segmentation results obtained by automatic methods. On the other hand, manual segmentation is a tedious and time-consuming procedure. It involves high workload due to intensive HCI and the quality of the results is prone to inter- and intra- observer variabilities [WHIT2013, HECK2013]. Semi-automatic methods are potentially the most promising approach [RAMK2016] as a well-designed semi-automatic method is able to combine the state of the art image analysis algorithm with physicians' expertise to contribute to the effectiveness and efficiency of the segmentation process.

1.4 Information flow in an Interactive segmentation

Semi-automatic segmentation (SAS) methods, which is also referred as interactive segmentation methods, are partially supervised automatic methods and they provide solutions by combining physicians' expertise and computerised medical image analysis [BOYK2001, YEO2011, LEE2008]. Figure 1.4 presents a typical information flow of an interactive segmentation method [STOR2011, OLAB2001]. The flow starts with a physician, who first perceives the information on the dataset to get familiarized with the case. After acquiring the information from the dataset, the physician decides on the next step in the segmentation process and performs the required action. Here the term action refers to the physical activities performed by the physician such as moving his or her hand to choose the input device, scrolling the mouse button to select the desired plane/tool, pressing the zoom-in/out button, initializing the segmentation by drawing contours, and positioning the hand in case of gesture interaction. Actions performed by the physician are interpreted by software via a graphical user interface (GUI). Once confirmed, the medical images are processed by a computational algorithm(s) utilizing the input(s), and the output data is displayed on the user interface. This process iterates until a satisfied result is achieved.

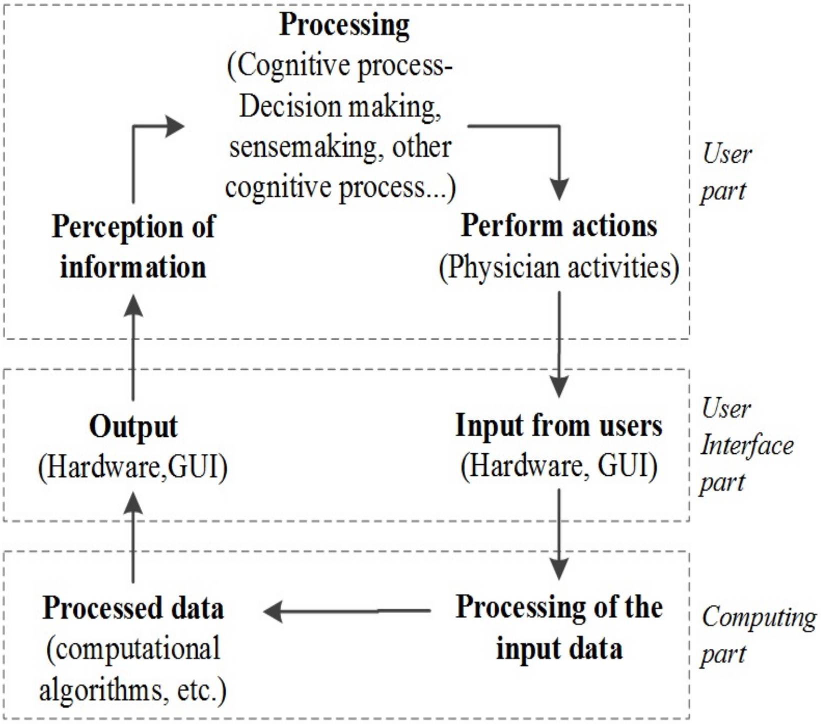


Figure 1.4: Information flow of HCI in an interactive segmentation method

1.5 Scientific gaps

Effectiveness and efficiency of an interactive segmentation method depend on the proper combination of physicians' expertise and the capability of the computational algorithms [KARR2008]. Though physicians play a crucial role in the segmentation process, most of the literature has focused on a specific aspect of the procedure regarding technical elements, such as testing the segmentation algorithm and system accuracy [BLAK2004, ZHOU2013]. For instances, McGuinness [MCGU2010] compared different interactive segmentation algorithms and evaluated the effectiveness of the system by measuring the performance and characteristics of the algorithms. The cognitive aspects of physicians and HCI in the segmentation process have been addressed in a few works [HARD2003, OLAB2001, YANG2010].

In the interactive segmentation process, various HCI components play important roles such as: 1) the user input approaches; 2) user input tools and 3) user input devices (UIDs). A poorly designed HCI in an interactive segmentation method may lead to higher workloads of the physician, may lead towards wrong inputs and may influence the quality of the results. With the advancement of new technologies, new applications of those HCI components can be found in different areas [ALON2013], especially regarding the user input approaches, tools and devices [LEAP2016]. However, only a few research works have been conducted to understand the impacts of those HCI components in medical image segmentation [MULT2011] and make improvements based on those understandings. The clear requirements on the user input approaches, tools and devices, and impacts of the HCI components on the segmentation results are still missing. For instance, Macchia [LAMA2012] evaluated three different pieces of radiotherapy segmentation software, but only evaluated the segmentation outcomes. The influence of the HCI process on the results was not discussed. Hornbæk [HORN2006] concluded that identifying relations between the HCI process and the outcomes from the measurement is an important aspect in a HCI research.

1.6 Research goal

The goal of this thesis is to propose effective and efficient HCI designs for the interactive segmentation. For this, concrete design requirements are needed regarding: What type of issues are physicians facing with the current systems? What types of information do they miss? What types of HCI are more effective? And what types of HCI are more efficient? This thesis will focus on answering the following research questions first:

1. What is the interactive segmentation workflow in using current commercial radiotherapy segmentation systems in RT planning? (Chapter 3)
2. What are the HCI and the design issues of current systems? (Chapter 3)
3. What kinds of user input approaches are preferred by the user in an interactive segmentation method and why? (Chapter 2)
4. What types of user input tools are needed to support those inputs in an interactive segmentation method? (Chapter 2)
5. What are the potential benefits and disadvantages of different input devices to support those interactions in the fields of radiology and radiotherapy? (Chapter 2)
6. What are the various evaluation methods and what are their benefits and disadvantages, respectively? (Chapter 2 and 3)

Based on the acquired knowledge, user input approaches, tools and devices will be developed/selected. The designed HCI components will be able to answer the final two research questions:

7. What are the impacts of different user input approaches and tools on the interactive segmentation process and result? (Chapter 5)

8. What is the impact of using different input devices on the interactive segmentation process and the result? (Chapter 6)

1.7 The approach

This dissertation addresses several research and design challenges. Examples of these challenges include assessing the interactive segmentation workflow complications, finding the current segmentation system issues, understanding physicians' "real" requirements and understanding possibilities of new user input approaches, tools and devices. For this, a series of design approaches and research methods are adopted in this thesis.

The proposed design research is a multidisciplinary research where multiple stakeholders are engaged and they are not familiar with each other's discipline. Therefore, to accelerate the design process, an iterative process of co-design research was applied during each of the phases. The co-design approach is based on the process described by Freudenthal et al. [FREU2011]. It is applied in order to: a) combine the theory and practice through reflection and modification during each cycle of activities and b) maximize innovation in the development of an effective and efficient HCI. Applying co-design also means that during the different activities of the iterative process, the author works within a multidisciplinary team which contains designers, oncologists, medical physicists, and computer engineers. The team is committed to collaborate within the workgroup [KLEI2003; DANE2006; FREU2010]. In the design process, every team member brings in new expertise to contribute to the solution [KVAN2000], which will support the production of a complete design [FREU2010]. During this research project, co-design was conducted by having frequent observations, discussions and brainstorming sessions, and by developing and testing prototypes within the team and with externally invited users. The collaboration among different stakeholders gives the opportunity to quickly, even on-site, fill the knowledge gaps, solve problems and verify design proposals. Methods from user-centred design, user interface design, physical ergonomics, and cognitive ergonomics or human factor (HF) [FREU2010] were also used within the approach. By combining the different research methods, the advantages of each could be utilised. Using the co-design methods, the proposed research can be divided into three phases: the exploration phase, the concept and design phase and the validation phase. Figure 1.5 illustrates these three research phases and the design process:

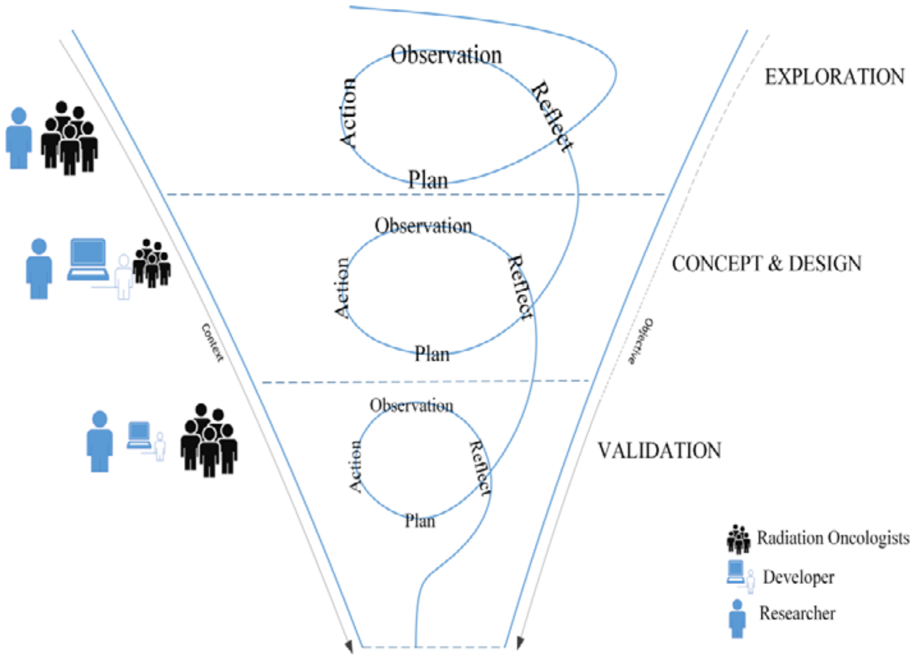


Figure 1.5: Three research phases and the iterative design process

- In the exploration phase, we studied different contouring tasks in order to systematically reveal different aspects regarding the design, such as the workflow, HCI engaged in the workflow and UI requirements. This was done by: 1) literature review; 2) studying the procedure, physicians' and their context using observational studies and the think aloud method; 3) understanding the existing commercial contouring systems using the heuristic and think aloud methods. The goals were to:
  - understand the workflow of different radiotherapy segmentation systems;
  - discover possible usability and HCI design issues of current segmentation systems;

    explore the abilities and limitations of various HCI evaluation methods.

- During the concept and design phase, we set up the design focus and iteratively designed creative solutions through: 1) brainstorming sessions; 2) workshops; 3) discussions and 4) mock-up testing. The goal of this phase was to:
  - setup the design focus;
  - conceptualize the designed user input approaches, tools and devices and iteratively verify the concepts in the co-design process;

- In the validation phase, we implemented the concepts and confirmed the findings by evaluating the designed HCIs with radiation oncologists in a realistic setting. This was done to:

- o understand the impacts of different user input approaches and tools;
- o understand the impacts of different user input devices;
- o validate the findings based on combined evaluations on the HCI process and the results;

Each research phase was an action cycle of four steps: 1) planning a change; 2) acting to realize the change; 3) observing the process and the consequences of the change and 4) reflecting on the process and its consequences [KEMM2014]. Other research methods are adopted as well if needed. Figure 1.5 also shows the involvement of various team members in different phases. As a result, requirements for effective and efficient HCI designs could be unveiled and at the same time, working prototypes were developed.

1.8 The Team

The proposed research was a part of the "SUMMER" project (Marie Curie Research Training Network (PITN-GA-2011-290148), 7th Framework Programme of the European Commission). "SUMMER" was created to support the technological and clinical research required for the innovative use of multimodal images in radiotherapy treatments. "SUMMER" aims to:

- Produce a new generation of software solution using all imaging techniques for biological target volume delineation, based on spatial co-registration of multi-modal morphological and functional images. Included imaging techniques are fMRI, MRS, 4D PET,
- Contribute to clinical efforts on better accuracy on target while increasing the safety for organs at risk.

The project was conducted by a multi-disciplinary team which formed at the beginning of the project. Figure 1.6 shows the partners of the SUMMER project. The team consisted of three hospitals (Universitätsklinikum-Frieburg, Fondazione Santa Lucia-Rome and Institut Claudius Regaud-Toulouse), two industries (Aquilab-Lille and VRVis-Vienna) and two universities (Medical university of Vienna and Delft University of Technology). Each member had a specific expertise and a different tasks. The author worked closely with Aquilab- the industrial partner, Universitats klinikum, Freiburg and Institut Claudius Regaud, Toulouse.

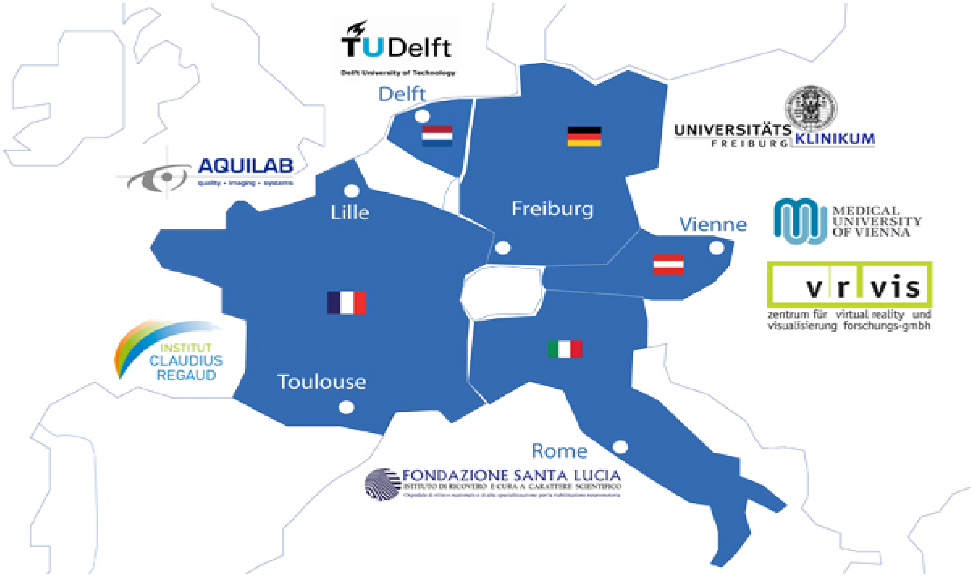


Figure 1.6 Partners of the SUMMER project

1.9 Thesis structure

Following the approach presented in Fig.1.5, this thesis reports the related activities in seven chapters. Figure 1.7 illustrates the logical relations among those chapters. Besides this chapter:

Chapter 2: Literature review reviews the relevant literature which includes different types of interactive segmentation workflows, user input approaches, tools, and the input devices used for interactive image segmentation tasks. Based on the reviews, the desired supports for the user during interactive segmentation were identified. Besides, various HCI evaluation methods, which are able to provide feedbacks on the designed user input approaches, tools or devices, are discussed as well; (Phase 1 and 2). The research questions that will be answered in this chapter are: What kinds of user input approaches are preferred by the user in an interactive segmentation method? What types of user input tools are needed to support those inputs in an interactive segmentation method? What are the potential benefits and disadvantages of different input devices to support those interactions in the fields of radiology and radiotherapy? What are the various HCI evaluation methods and what are their benefits and disadvantages, respectively?

Chapter 3: Field Research aims at identifying workflows of existing software solutions and finding the usability and HCI issues of those software solutions regarding the functions, user satisfactions, limitations, frequently encountered human errors and workloads. The following research questions will be answered in this chapter. What is the interactive segmentation workflow of using current commercial radiotherapy segmentation systems in RT planning? What are the HCI and the design issues of current systems? What are the various evaluation methods that can help in identifying the design issues? In order to answer the research questions, three software solutions were evaluated in different hospitals. The observational research methods, the heuristic evaluation method, the think-aloud and NASA task load index(NASA-TLX) questionnaires are used to get an overview of the workflow and the HCI process of current segmentation systems that are being used in various hospitals for contouring. The obtained insights will be helpful to propose new design requirements which might fit the wish of radiotherapy physicians; (Phase 1 and 2)

Chapter 4: Prioritization and Design focus integrates the knowledge from the previous chapters and presents the strengthened framework of primary improvements needed; (Phase 2)

Chapter 5: User input approaches and tools present the designed user input approaches and tools for interactive segmentation. In this chapter, we will answer the following research question: What are the impacts of different user input approaches and tools on the interactive segmentation process and the result? For this, two different approaches which utilize different user input tools are being compared to investigate their effectiveness and efficiency on delineating organs at risk regarding the HCI process and the results; (Phase 2 and 3)

Chapter 6: User Input devices presents the design and the evaluation of an interactive segmentation method which utilize four different HCI input devices, i.e., the mouse, the pen on pad, the pen on screen and the touch screen. The design is developed based on the insights from *Chapter 4* and the research question that will be answered in this chapter will be: What are the impacts of different input devices on the interactive segmentation process and the result? 12 radiation oncologists participated in the experiments to evaluate the impact of the input devices on the image segmentation process and the results, respectively; (Phase 2 and 3)

Chapter 7: Discussion and Conclusion discuss the outcomes of the research. Limitations of the research are presented as well to provide suggestions for the future research. Finally, the original contributions of this research are summarized. (Phase 3)

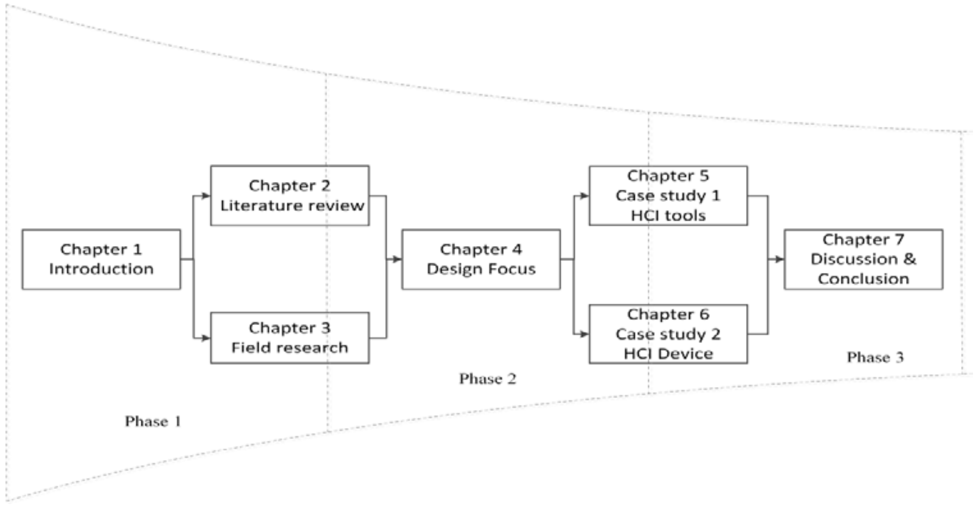
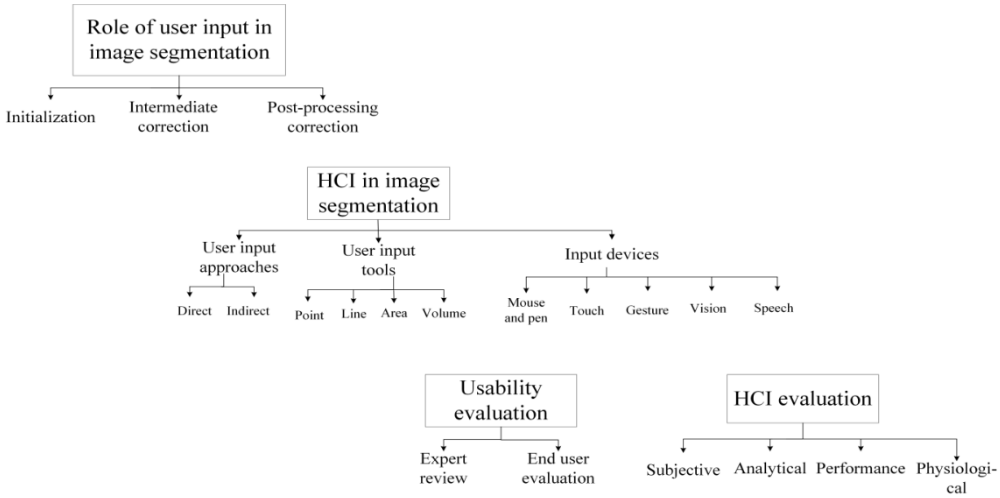


Figure 1.7: The structure of this dissertation

2 Literature Review



This chapter is based on

Anjana Ramkumar, Pieter Jan Stappers, Wiro J.Niessen, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle and Yu Song
"Using GOMS and NASA-TLX to evaluate Human Computer Interaction process in interactive segmentation". International Journal of Human-Computer Interaction, Vol 33(2), 123-134, 2017 .

Anjana Ramkumar, Yu Song, Edit Varga, Wiro J Niessen, Anne Laprie, Ben Rowland, Adinda Freudenthal.
"Comparison of Heuristic Evaluation and Think Aloud Methods A Study in Radiotherapy Contouring Software". Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare 2014 June; 3(1): 230-237.

Anjana Ramkumar, Edit Varga, Wiro J Niessen, Adinda Freundethal.
"Exploring input devices for contouring in external radiotherapy". Innovative imaging to improve radiotherapy treatments, Lulu Enterprises Inc Ed, 1:29-34, July 2013.

This chapter presents a review of existing literature to support the study undertaken in this thesis. Google scholar (www.scholar.google.com), Scopus(www.scopus.com) and Science Direct (www.sciencedirect.com) were the search engines used to obtain relevant literature. The search for articles were based on the following query words - ‘‘user interaction”, “human-computer interaction”, “medical images”, “semi-automatic segmentation”, “interactive segmentation”, segmentation”, and “contouring”. These words were used in different combinations to search within abstracts and title. The search was limited to published works between 2008 and 2016. It was decided to focus the review more on the organ segmentation and less on the tumour segmentation. It is true that there are many other fields which use segmentation. However, other images might not be as complex as medical images. Based on the literature search, we have categorized this chapter as four sections: Section 2.1 identifies the role of user input that is required during interactive segmentation. The elements of HCI that are involved in interactive segmentation are identified in Section 2.2 and are categorized as: Section 2.2.1 describes user input approach; Section 2.2.2 summarizes user HCI input tools and Section 2.2.3 investigates HCI input devices. Section 2.3 reviews the methods and measures that are used in usability evaluation and finally Section 2.4 reviews the HCI evaluation. The conclusions of this chapter are presented in Section 2.5.

2.1 Role of user input in image segmentation

Interactive segmentation [OLAB1997] plays an important role in the segmentation of medical images, where user involvement is considered as a supplement to the computational algorithms. This technique leverages the expert knowledge of users, which facilitates accurate segmentation and the treatment of various tumours. From the literature review it was identified that, users are involved in different stages of the interactive segmentation workflow: Initialization, intermediate correction and post-processing correction.

User interaction during Initialization: In the initialization process, users are required to give hints to the algorithm regarding the location of the ROI [DOLZ2014a, EGGE2014]. Figure 2.1 shows an example workflow of the user initialization when the image data is available within the software. In the figure, after the image data is presented to the user, he/she initializes the data using the input tools in a single slice. Once the system gets the hint in one slice, the algorithm runs over for all the slices to calculate the results which will be presented to the user.Generally, the more accurate the hints are, the better the outcomes are, which

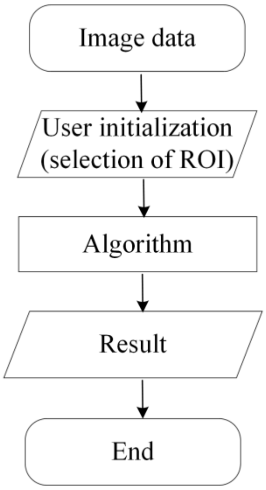


Figure 2.1: An example of the process of Initialization

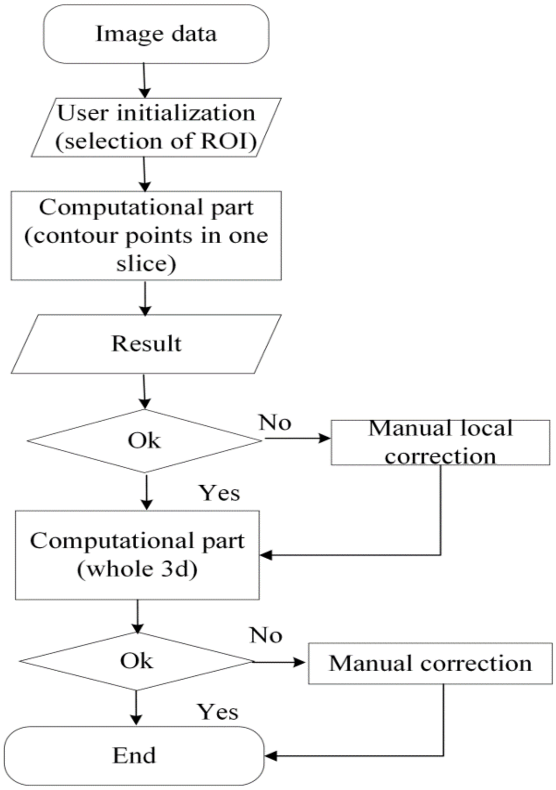


Figure 2.2: An example of Intermediate correction

further reduce the workload of the users. If the user is satisfied, he/she can accept it by saving the result. In some cases, user initialization was coupled with the insertion of some pre-defined templates of shape [EGGE2014]. For example: insertion of organs shape, rectangle, square, etc.

User interaction during Intermediate correction: Figure 2.2 shows an example of the intermediate correction process. There are many different ways to do it such as predictive modelling, etc. In the example shown, the role of user in this phase is to revise the outcome from the initialization process and to make the corrections in only one slice [ZHOU2013, KOCK2014]. With the corrected slice as an input, the algorithm revises the contour in rest of the slices. From Fig.2.2 it can be seen that manual local correction was required in the intermediate step. If the user is not satisfied with the results, he/she can reinitiate manual local corrections, then the algorithm is computed again based on the corrections. Or if the results are still not satisfactory, the user can perform manual corrections slice by slice.

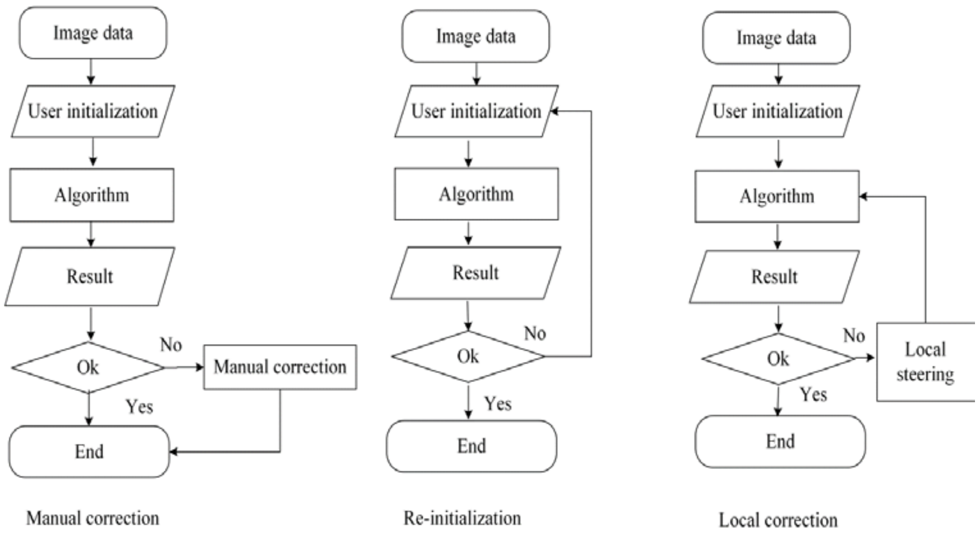


Figure 2.3: Examples of post-processing Interactions

User interaction during Post-processing correction: Post- processing correction is the process of correcting the results generated by an algorithm. Three types of post processing corrections are typically being used in the workflow of medical image segmentation as Fig. 2.3: 1) Manual correction - the process of correcting the segmentation outcome manually slice by slice until a satisfactory result is achieved [DVOR2014] as the left diagram in Fig.2.3, where the users have to correct their outcomes manually if they are not satisfied with the results of the initialization process ; 2) Re-initialization - the user re-initializes the segmentation outcome using the different initial inputs [ZHU2009] as the centre of Fig.2.3, where the users initialize their outcomes again, if they are not satisfied with the outcome. If they are satisfied with the first initialization, they save the contours and finish the segmentation process. Re-initialization usually modifies the whole initial outcome and hence might result in a new segmentation outcome; 3) Local post-processing correction - The role of the user in this method will be to identify and select the “incorrect” region for post processing correction as the right Fig.2.3. Using this method, if the user is not satisfied with the outcome, he/she would select the particular region which is incorrect. After selecting the region, the user needs to modify contours in that region, for instance, by the same computational algorithm but only applied in this region [BEIC2012, HECK2009].

2.2 HCI in image segmentation

Human-Computer Interaction (HCI) lies at the crossroad of several research areas including computer vision, psychology, etc. [JAIM2007]. Computers are as ubiquitous in healthcare as in modern society [FROM2011]. HCI plays an important role among all healthcare professionals and clerical staffs to keep the track and view of patient records, for making appointments, etc. [FROM2011]. The main expectations of HCI in healthcare are user friendliness, user-acceptance and user-competence [LUN1995]. In order to achieve efficient HCI in interactive segmentation, various HCI components such as: 1) user input approach, 2) user input tools and 3) user input devices (UIDs) play important roles. The main goal of this section is to explore the potential benefits of different HCI components. This will help to identify the possibilities of improving user experience in the field of radiotherapy contouring.

2.2.1 User input approach

Two types of user input approaches are often used in contouring: 1) the direct approach and 2) the indirect approach. In the direct approach, the user will directly specify (part of) the outputs of the delineation task. For instance, using the pictorial inputs where the pixels indicated by the user serve as resulting segmentation, usually leading to low-level determining the image properties of the object by hand. The direct approach is the most popular approach in current segmentation software solutions, for instance, the live wire and live lane [ZEWE2014], intelligent scissors [MISH2008], etc.

In an indirect approach the user needs to roughly specify the locations of the organs, then the computational algorithms compute the output based on these “hints” given by the user. Some examples of indirect approach are: Atlas[ISAM2008], graph-cut[DOLZ2014a], grab cut[ROTH2004], etc. In the atlas based segmentation, the knowledge about the shape, object orientation, continuity, elasticity or smoothness of the object are incorporated into the system. This prior knowledge is extracted from a reference image which is often called atlas. Even though the user does not have to specify anything directly on the image, the system will be able to choose the right atlas from the database. Hence it is an indirect approach which will have an influence on the result. It is worth mentioning that the effectiveness and efficiency of either the direct or the indirect approach strongly depend on the incorporated segmentation algorithms, if any. As this thesis will not focus on the computational algorithms in the implementation, a combination of graph-cut and watershed-based algorithms which was able to facilitate both approaches [DOLZ2014a, DOLZ2014b], was adopted to test the different HCI approaches.

2.2.2 User input tools

In medical image segmentation, most of the research has focused on the algorithms of the tools than on the HCI process of the tool [KANG2004, HECK2013]. Olabarriaga et al. [OLAB2001] investigated HCI issues in 2D segmentation and they found that deform, edit boundary, and rectangle are some of the most frequent tools used in segmentation. Aselmaa et al. [ASEL2013] concluded that in manual segmentation; brush tool, 3D pencil, smart brush, and nudging were often used. Zhao et al. [ZHAO2013] classified the user interactions in segmentation of medical images into three: menu option selection, pictorial input on an image grid, and parameter tuning. Among the three approaches of user interactions, menu option selection is considered as the most efficient way, but it constrains the freedom of user's choices. The pictorial input is simple, but it could be time-consuming. Parameter tuning is easy to operate, but it may require specific training for insights of the automatic computational part. The pictorial input can be further categorized in four categories: the point based inputs, the line based inputs, the area based inputs and the volume based inputs as Fig.2.4.

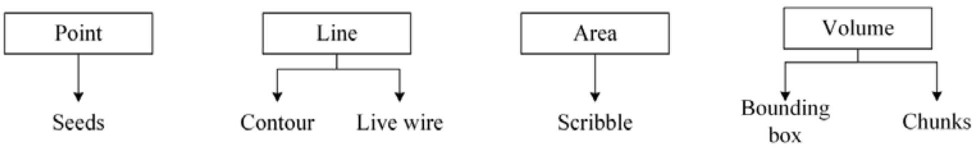


Figure 2.4: Types of pictorial inputs

The *point* based inputs use one or more individual points as the input for segmentation. These single points are often named seed and they are usually put in the centres of the ROIs [VELA2013, STEG2012]. This is typically an initialization step in order to give the algorithm an initial guess about the location of the ROI to be segmented. Additionally, the seed points can also be used for manual post-processing [SUN2013, KARA2013] in order to refine the result by inputting additional seed points. The advantage of the seed points based input is that it requires few user interactions.

The next type of user input is the *line* based input, for instance, the user can use it to trace the boundary of the organ by a mouse or a pen input device. This line is most commonly known as the contour in medical image segmentation. Another form of line based user input is the live wire [BARR1996, LI2012, WANG2012] where the line drawn by the user is being changed by the algorithm to match the extracted edges. When the mouse position comes in proximity to an object edge, a “live wire” boundary snaps to it, and wraps around the object of interest. The line based interactions are often used during initialization and post-processing corrections. For instance, Heckel et al. [HECK2009] used the live wire input for post-processing local correction. The line based input might be time consuming, as well as physically and cognitively challenging.

The third type of input is the *area* based input, where the user needs to indicate an area of interest. In many studies scribbles are used as an area based inputs for initialization [GAO2012]. Scribbles are combinations of multiple seed points. With the scribble input, users will be able to adjust the width of the scribble (e.g. narrow or wide) depending on the size of the ROI.

The fourth form of user input is the *volume* based input. Volume based inputs could be bounding box or segmentation chunks. With a bounding box, the user could select the ROI that they want to segment in 3D. Depending on the size of the region, the bounding box size can be adjusted to the users’ needs. The bounding box inputs are used either during intermediate correction or during post-processing corrections. Chunks based input requires the user to specify the volume chunk that is incorrect in the segmented image. Beichel et al. [BEIC2012] had used a chunk based interaction for post processing correction, where the correction is made by removing chunks of contour from the image.

2.2.3 Input devices

In the workflow of interactive segmentation, mouse, keyboard and monitor screen are often used as input devices to achieve desired HCIs. However, many others devices may facilitate this process as well.

Mouse and Pen interaction

In clinical practices, mouse and keyboard are common HCI devices which are used in medical image segmentation. Using them requires few training and generally, the outcome is satisfactory. However, in radiotherapy planning, the advancements of technology in the past decades have made it possible to deliver the radiation to very complex shapes [NUTT2000]. Therefore, higher accuracy is needed in identifying tumorous tissues and OAR for a better outcome of the treatment. Besides, the user often uses mouse and keyboard in other tasks. The overuse of these devices may lead to increasing amounts of repetitive strain injuries (RSI) [DAMA2000]. Thus, a new form of interaction which would help user to improve the accuracy, speed up the segmentation task, minimize risk of wrist injury when processing large amounts of data is needed.

In the past decade, the pen was introduced to segmentation task. The reliability, accuracy, and user satisfaction of using both the mouse and the pen have been tested in medical image segmentation [CHEN2011, PERA2011]. It was found that the performance of using a pen was generally better than the mouse and the overall error and the time taken for segmentation was less as well. Regarding the muscular load, Kentaro and Horii [KOTA2003] discovered that the performance of the pen exceeded the performance of the mouse. Sherbondy et al. [SHER2005] evaluated performing a simulated angiography localization task with the trackball, the pen, the jog-shuttle wheel and the mouse, respectively. They found the pen input devices in two distinct configurations performed faster than the mouse and trackball.

Touch based interaction

With the development of technologies, touch screens are gradually introduced to the medical context [MCWA2005], and touch-controlled interfaces were provided to many applications for image review [BAUM2011, SZÉK2013]. There are many benefits of using touch screen for decision making tasks and collaborative works, for instance, per-operative planning [LUND2011]. Generally, a touch interface is intuitive [LUND2011], and it is found to be more efficient than the mouse for selecting and sorting tasks [KIN2009]. Even though a bigger screen has a larger view area, physicians are more in favour of smaller sized tablets, as they are portable and the physicians can accomplish tasks anywhere. Regarding medical image segmentation, based on a cartilage segmentation task, McWalter et al. compared the segmentation time, precision (reproducibility) and measurement consistency of three input devices: mouse, digitizing tablet, and touch sensitive screen [MCWA2005]. They identified that segmenting with an interactive touch screen reduced segmentation time by 15% when compared to the traditional mouse but no significant difference was identified between the digitizing tablet and the traditional mouse.

Apart from healthcare multi-touch has been used in many other areas such as social, educational applications [KIM2016, HUNG2016]. Studies on the physical ergonomics in touch screen regarding comfort level, user preference/satisfactions, posture, muscle load, etc. reveals more advantages regarding its performance [PARK2010] and accuracy [PARK2010]. However, the effects of long term usage may differ. For instance, Bachynskyi et al. [BACH2015] compared the physical ergonomics of long term usage of the touch screen, the pen on tablet, the laptop, the table top, the public display and the smartphone. They identified that except the tablet and the laptop, other devices are not suitable for long term usage unless with proper posture. Even for the tablet, they mentioned that it is suitable for long term use only after adjustment of the posture to avoid neck problems.

Gesture-based interaction

Using computer keyboards and mice in intensive care units (ICUs) may spread infections [SCHU2003]. A gesture based input can play an important role in such situations for interaction with computing devices. With a wave of a hand or lift of a finger it is possible to change the way we interact with computers. The gesture-based interaction approach can be divided into glove-based method and vision-based approach. With the advancement of technology, the glove-based approach has largely been replaced by vision-based techniques. Vision-based hand gesture interaction has already shown to be a rapid and intuitive interaction approach in brain biopsy procedures for navigation and manipulation of images [WACH2008]. Rautaray [RAUT2012] has published a survey on hand gesture recognition in HCI, where they have also described about gesture based interactions used during laparoscopy and other surgeries. Many people have developed a Kinect-based intra-operative medical image viewer for use in a surgical environment [BIGD2012, GALL2011]. With such system a doctor could manipulate a medical image without touch the system during a surgery, such as zooming in, moving the image around, add a label at the specific place in the image. Gesture based interaction are not so popular in medical image segmentation. Chang [CHAN2016] used a 3D image interaction system and an image segmentation process based on gestures and voice commands through the Kinect sensor. Such a combination of interaction improved their interaction efficiency and reduce leakages in the segmentation refinement process.

New technology is emerging in the market, known as the Leap Motion Controller [LEAP2016], which has been claimed to be more accurate than the Kinect technology. One of the main differences between Kinect and Leap Motion is that with the former the user needs to be far from the screen whereas with the latter the user needs to be very close. The Leap Motion allows to manipulate the screen via a series of hand and finger movements in the air. Moreover, it is not only able to detect finger movements, but also a pen or a pencil could be used for drawing. Distance, however, remains a limiting factor for both technologies, as user have to operate at a certain distance from the screen.

Vision based input

Besides those devices, vision based inputs, which have been used in other fields like air traffic control [ALON2013, JAUM2014, LUPU2013], etc., were introduced to medical image segmentation as well. For instance, Sadeghi et al. [SADE2009] investigated the possibility of using eye gaze to perform segmentation task. They found that accurate placement of strokes might be strenuous on the eyes for complicated medical images. Noronha [NORO2013] identified that using eye gaze as the input device, the frustration level of the user is the same as using conventional input devices such as mouse.

Speech/voice interaction

Physicians are generally reluctant to use interfaces that require a considerable amount of typing [SHIF1991]. Speech recognition technology has the advantage that it does not tie the user down to the keyboard and it provides some freedom to interact with the system even at some distance away [LUN1995]. The user can enter or receive data while engaged in other work. Similarly, the doctor can examine a patient and could record his findings using voice input [LUN1995]. There are many software solutions existing where the report can be directly dictated into the computer system [HÖTK2013]. Furthermore, it could also be used to control system tasks [SHIF1991]. In medical image segmentation, speech interaction is just upcoming. Gering [GERI2016] introduced voice activated image segmentation which allowed the physician to quickly and easily interact with the computerized segmentation process, thereby imparting his/her skilled expertise to the segmented result. Speech interaction in combination with gesture interaction is very useful in a sterile environment for interventional radiology procedures [HÖTK2013]. Even though this interaction has many benefits, the negative side of it is that speech interaction needs a quiet environment and continuous speaking, which may lead to fatigue at some point [HÖTK2013].

In summary, a number of input devices which are able to help physicians in medical image segmentation are available. However, a better device in other applications need not necessarily outperform others in the segmentation as it requires high accuracy, high efficiency and long term comforts simultaneously. For instance, Molin et al. [MOLI2015] compared mouse, 6 degree of freedom controller and a touchpad for digital slide navigation for the pathologist. They identified that participants perceived less workload with the mouse. A better understanding of the effectiveness and efficiency of using those devices is a prerequisite in selecting the proper device and making suggestions on the interface.

2.3 Usability engineering evaluation

Traditionally, HCI bridges psychology and informatics, while Usability Engineering (UE), as an engineering discipline, is anchored in software technology thus enabling appropriate technological implementation. Together, HCI and UE provide an emerging potential to assist the daily workflows in the realm of medicine and health care. Usability is a quality attribute that assesses how easy user interfaces are to use. The word "usability" also refers to the methods for improving ease-of-use during the design process. ISO 9241 part 11 defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [ISO1998]. Here effectiveness refers to how completely and accurately the work/goal is reached. Efficient refers to how much effort, time, and/or costs users paid to finish a task. Satisfaction denotes how much users are satisfied with the process of completing the given task. The expert review methods involve usability specialists who systematically inspect the software platforms and then identify best practice design violations. These reviews are typically conducted in order to classify and prioritize usability problem. Observations [MILL2000], heuristic [ZHAN2003], cognitive walkthrough [WHAR1994] etc., are typical expert review methods. But the scope of this thesis is narrowed to observational and heuristic method as the former one gives a broad understanding of the whole scenario and the latter one is a quick and cheap in evaluating user interfaces.

Observational methods have become increasingly popular in the field of UE and HCI [MILL2000]. This method describes the behaviour, communication patterns, workflows and tasks of users in specific work environments. Rose et al. [ROSE1995] used the observational method to re-designing a user interface and they concluded that observational methods based on principles of participatory design have proven to be an effective tool in user interface redesign. Chan et al. [CHAN2012] have used the workflow analysis and observational method to investigate the potential errors with the radiotherapy treatment planning software. Monahan et al. mentioned that the advantage of the observational method is that it is conducted in real world context and generate a rich data directly from the users

[MONA2008]. However, they also mentioned that observational studies are labour intensive, studies tend to have a long timescale, and there may be difficulties with data analysis.

Heuristic evaluation is a typical expert review methods for quick, cheap and easy evaluation of user interfaces. It is often carried out by a small set of evaluators examining a user interface who judge its compliance with a set of recognized usability principles. The initial set of heuristic rules was developed by Nielsen and Molich in 1990 [NIEL1990]. Later Zhang et al [ZHAN2003] further developed the methods by adding few more heuristics to the Nielsen's rules. Lilholt et al. [LILH2015] used heuristic evaluation to evaluate a telehealth system. They concluded that heuristic evaluation was an effective method for uncovering and identifying problems with the system. The consistent finding of particular usability problems confirms that the development of a telehealth system should pay particular attention to user aspects. Besides, they suggested that heuristic evaluation can always be followed by user tests to evaluate the design of telehealth systems. In radiotherapy, heuristic evaluation has been used to evaluate safety of medical devices by identifying usability issues. For instance, to explore possible improvements of the interface and HCI of those systems, Chan et al. conducted heuristic evaluation of a user interface of a software system developed for radiotherapy [CHAN2012].

End users, such as clinicians and allied health professionals who will actually use the system in the real environment, may participate in usability testing in which they are given tasks to complete. And they are able to report their experiences and opinions. Many methods are available such as interviews, questionnaires and the think aloud method. In this section, we focus on interviews and the think aloud method, the questionnaires will be discussed in the HCI evaluation section.

The think aloud method had been often used in the healthcare domain [JASP2009]. This method was introduced by Clayton Lewis in 1982, in the field of usability [LEWI1982]. In think aloud method, participants express their thoughts on the application while executing set tasks. Using this method, the designer can gain deep insight of the problems that end users encounter in interacting with a system. It will also lead to a better understanding about the cognitive processes of the users and therefore to building user interfaces on the basis of these insights. Jaspers. [JASP2009] used think aloud method in combination with video recording to get a deep understanding of the way in which four paediatric oncologists searched through the paper-based patient records in preparing a patient visit. They concluded that cognitive engineering methods in the system design process may be of great help in designing systems that fully support health care professionals’ work practices. Kelly [KELL2009] has identified that the protocols of the think aloud method may be cognitively complex for users. But there is some evidences that if proper training may improve the data collection process.

Interviews are another way of exploring the user experience during usability evaluations. Interviewing is a commonly used technique where users, domain experts and/or other stakeholders are asked questions by an interviewer in order to gain information about their needs or requirements. Usually interviews are structured or semi-structured interviews. Structured interviews are usually carried out in situations where the respondents' range of replies is already well known and there is a need to gauge the strength of each shade of opinion. Semi-structured interviewing is useful in situations where broad issues may be understood, but the range of respondents' reactions to these issues is not fully known. Post-task interviews can be used to probe more deeply on interesting issues. The post-task interview allows observation and verbalization data to be obtained quickly without analysing tapes. Post-task interviews can offer benefits at the cost of slightly longer evaluation sessions with children. For instance, Vermeeren et al. [VERM2007] conducted a study on the use of post task interviewing evaluation technique with 6-8 years old children. The results showed that children overall were fairly good at answering the questions. Though interviews have many advantages, if the questions are not prepared and asked properly, then the interviewer might not get the complete answer. And this can lead to incomplete and misleading conclusions.

Many authors have compared various usability evaluation methods in healthcare and other fields. Jasper compared two expert based methods (Heuristic and cognitive walkthrough) and one user based method (Think aloud) [JASP2009] for testing interactive health technologies. His study concluded that a combination of different techniques that complement one another should preferably be used as their collective application will be more powerful than applied in isolation. Similar result was identified by Yen et al. [YEN2009], where they used heuristic and think aloud to evaluate a web-based questionnaire for nurse scheduling.

2.4 HCI evaluation

Experimental methods are often used in HCI evaluation. Depending on different applications, the setup of the experiment may vary. However, four type of measures, named subjective measures, performance measures, physiological measures and analytical measures as summarized by [GAO2015] are the most commonly used measures in HCI evaluation. In this section, we will be focusing on the aforementioned measures.

Subjective measures

Subjective measures are designed to collect the opinions from the user about the workload/human effort, satisfaction, preference, user-experience, etc. In spite of the criticism on the validity and vulnerability to personal bias of the self-report method, subjective measures, with the low cost and the ease of administration, as well as adaptability, have been found highly useful in a variety of domains, including healthcare, aviation, driving and even office working environment [LONG2011, ROSC1990; MORG2011; BRID2011]. The most common way of obtaining the subjective measure is through questionnaires. Terwin et al. [TERW2015] identified that Likert-type items are widely used in human-computer interaction research to measure subjective user experience [LAZA2010, KAPT2010, NIEL1994]. Established usability measures such as ISO-9241-9 [ISO2000], SUS [BROO2013] and QUEST [DEME1996] use Likert-type items. They are appropriate for smaller sample sizes, easy to learn and quick to execute. They can be presented verbally, on paper, or digitally.

Subjective measures can be used to measure the cognitive workload. NASA-TLX [HART1988] is one of the most widely used instruments and has been extensively tested in human factors studies for the measurement of different types of workloads. NASA-TLX consists of a set of six rating scales to use in evaluating the workload of the physicians in a task. The six rating scales are "mental demand, physical demand, temporal demand, performance, effort and frustration". Each rating scale is divided into 21 gradations starting from 0 to 100. The comparisons of sensitivity and diagnosticity between NASA-TLX and other subjective measures have been a long and on-going debate, NASA-TLX consistently exhibits high reliability, user acceptance and low inter-subject variability in various research [RUBI2004; DEY2010; CAIN2007]. In HCI NASA-TLX has been used to identify users emotions, metal demands, performance etc. [JEON2015; GAO2015] In radiotherapy, several studies have been using the NASA-TLX to identify physicians' workload during various stages of the workflow [MAZU2014, MOSA2011; RAMK2015].

Analytical measures

Analytical evaluation methods are popular in HCI evaluation because they often require less formal training, take little time to perform, and can be used in both early and late stages of the development process. Models that quantify estimated workloads were often used in analytical evaluation. Previous research indicates that using models are more consistent and quantifiable than using individual measures. However, it should also be noted that accuracy of the model highly depends on the completion of the tasks and the time required for building such a model also depends on the complexity of the task. A classic example of the analytical model is the GOMS [CARD1983]. GOMS is a specialized human information processor model for HCI observation. It is a method for describing a task and the user's knowledge of how to perform the task in terms of goals, operators, methods, and selection rules. Here *Goals* refers to a particular state the user wants to achieve in their software or service. A user's goal can usually be divided into *sub-goals*, which may in turn need to be divided into even smaller sub-goals. Thus, the user's goal and sub-goals form a hierarchy. Goals are achieved by methods, which themselves contain operators that must be performed in a particular sequence to accomplish that goal. If there are multiple methods to accomplish a goal, selection rules are listed. *Methods* are well-learned procedures for accomplishing the goals. A method consists sequences of steps for accomplishing the goal. The classic example of a method is "deleting a paragraph in a text editor": Using a mouse, place the cursor at the beginning of the paragraph, hold the mouse button down, drag to the end of the paragraph, release, highlighting the paragraph, then hit the delete key. Another (less efficient) method can be: place the cursor at the end of the paragraph and hit the delete key until the paragraph is gone. *Selection rules* are used to determine which methods to select when there is more than one available for a given stage of a task. *Operators* are the actions that are performed during a process. With the original command-line interfaces, an operator was a command and its parameters, typed on a keyboard. In graphic user interfaces, typical Operators are menu selections, button presses, or user's actions. In some studies, gestures, spoken commands, or even eye movements are considered as Operators [LIN2013]. Operators can actually be defined at many different levels of abstraction.

In 1983, Card et al. [CARD1983] initiated the study of GOMS by their CMN GOMS model. CMN GOMS has a strict goal hierarchy and methods are represented in an informal form and can include sub-methods. Apart from CMN GOMS, many other types of GOMS models have been discussed in the literature: the Keystroke-Level Model (KLM GOMS) [KIER1993], the Natural GOMS Language (NGOMSL) model [KIER1988; KIER1997], the Cognitive Perceptual Motor (CPM-

GOMS model) [JOHN1996], and a more recent variation of GOMS named Sociotechnical GOMS (SGOMS) [WEST2007]. The KLM GOMS model is a simplified version of the CMN-GOMS model. It only utilizes six primitive operators as: 1) pressing a key; 2) moving the pointing device to a specific location; 3) pointer drag movements; 4) mental preparation; 5) moving hands to appropriate locations and 6) waiting for the computer to execute a command. A more rigorously defined version of the KLM GOMS model is named the NGOMSL model (KIER1988; KIER1997) which presents a procedure for identifying all the GOMS components, expressed in a form similar to an ordinary computer programming language. The NGOMSL model includes rules-of-thumb about how many steps can be part of a method, how goals are set and achieved, and what types of information should be remembered by the user while doing the task. The CPM-GOMS model was introduced to describe parallel activities [JOHN1996]. It utilizes cognitive, perceptual, and motor operators in a critical-path schedule chart to resemble multitasking behaviours of the user. West et al. [WEST2007] developed Sociotechnical GOMS (SGOMS) model, which extends the idea of using a control structure for dealing with processes such as planning, scheduling, and teamwork from micro to macro level tasks. SGOMS consists of two components: the first part of SGOMS is the planning unit which is a sequence of unit tasks for accomplishing a specific goal, the second component of SGOMS is a framework that describes how planning units fit into the work process. Christou et al. [CHRI2012] developed a new GOMS model named *codein* to support the evaluation of reality based interaction styles. The main advantage of their GOMS model was that it was able to evaluate the task completion time of parallel actions during the performance of a task which was only possible using CPM-GOMS.

In the past decade, GOMS model has been extensively applied in developing analytic models of user behaviour for user interaction evaluation. Carmel et al. [CARM1992] applied the GOMS model to analyse hypertext browsing strategies with a HyperCard application. They treated browsing as a cognitive information processing activity, and attempted to describe the browsing process both qualitatively and quantitatively. In their research, they identified three different types of browsing patterns: search-oriented, review and scan. In addition, they also compared tactics used by novice and expert users on a specific topic. Smelcer [SMEL1995] used a NGOMSL model to identify causes of user errors for database query composition. Saitwal [SAIT2010] also used the GOMS model to evaluate the electronic health record (EHR) systems and proposed suggestions for improving user interfaces. GOMS has also been successfully used to determine the usability of websites for disabled users [SCHR2006], to measure the performance on how users interact with web applications [ANDR2014], to assess the performance of automobile human–machine interfaces [XIAN2010], and the navigational structure of websites [OYEW2011]. Although it was designed to predict task execution time on mouse and keyboard systems, the GOMS model is flexible enough to be adjusted to measure the HCI performance of using touch screens [ABDU2011] as well.

Performance measure

Performance measures are based on the observable performance of the users while doing a task. Many studies have analysed the performance of input tools and devices [BACH2015] using different performance measures. In medical image segmentation, Kentaro et al. [KOTA2003] proposed that the process performance measure includes the duration of the process, the time to complete each segmentation task and the error rates. For measuring the performance of the HCI process, many studies have utilized video analysis and log files as tools [JAKO2016, SZÉK2013]. Besides, [DRUC2002] and [ROSC1990] also used a result-oriented performance measure by measuring the accuracy of the outcomes, respectively. In the area of medical image segmentation, the Dice-Jaccard coefficient (DSC) [DICE1945] and the direct Hausdorff distance [HAUS1962] are frequently used tool.

Physiological measures

Many physiological evaluation methods are intrusive, which may influence the behaviour of the user [DIRI2011]. Therefore, sufficient attention and time should be given to the user for training before using those measures. Most of the research pertaining to psychophysiological methods in these areas focuses on the mental workload assessment methodologies [KRAM1991, FARM2003, CAIN2007]. Dirican and Göktürk [DIRI2011] had identified various advantages and disadvantages with the physiological measures. According to their research, the six main advantages of those measures were objectivity, multidimensionality, unobtrusiveness, implicitness, continuity and responsiveness. They also summarized the disadvantages of the physiological measure as: special equipment is needed, data quality and interpretation are important, and unnaturalness of the use in the evaluation. In this section we discuss some of the most commonly used physiological measures in HCI studies.

Electroencephalography (EEG)

The EEG records the electrical brain signal from the scalp, where the signal originates from postsynaptic potentials, aggregates at the cortex, and transfers through the skull to the scalp. EEG based device that requires extracting raw EEG signals from the brain and converting it to device control commands through suitable signal processing techniques. Fery et al. [FERY2013] reviewed using EEG in the HCI context and identified that workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion as being recognizable by EEG signals. They also identified that workload, attention and emotion assessment benefits the most when an EEG evaluation method is used. A great challenge involved in using EEG relates to the presence of measuring artifacts, which originate from electrical impulses that are unrelated to cerebral activity. Such artifacts may originate from muscle tension, heart beats, eye blinks or body movement of any kind. Furthermore, the EEG may pick up signals from electronic equipment in the test environment. Also the EEG data are complex waveforms that require sophisticated signal processing equipment. Chanel et al. [CHAN2009] mentioned that, most contemporary EEG systems were equipped with robust software, which may facilitate data analysis by removing some of the most common artifacts. Also it is worth mentioning that most EEG measurement devices are intrusive device. Hence, in most of the studies, EEG measurement is not used as a stand-alone method, but combined other evaluation measures [BELL2010].

Electromyogram (EMG)

EMG measures electrical currents that are generated in a muscle during its contraction and represent neuromuscular activities. EMG signals can be used for a variety of applications including clinical applications, HCI and interactive computer gaming. For instance, Lozano et al. used EMG measure in evaluating interactions with a multi-touch tablet and identified that multi-touch interactions can induce significant stress that may lead to musculoskeletal disorders [LOZA2011]. Kotani obtained EMG measures of four muscles (descending part of upper trapezius, biceps brachii, flexor digitorum superficialis, and extensor digitorum) to evaluate muscular load of pen-tablet versus mouse [KOTA2003]. Hazlett used facial EMG to measure the continuous emotional state of the user when evaluating the usability of two websites [HAZL2003]. They mentioned that moment to moment emotional experience of the user can be related to interface events, navigation logs and other user behaviour to give a dynamic understanding of the human computer interaction. Compared to other bio-signals, EMG contains complicated types of noise that are caused by, for example, inherent equipment noise, electromagnetic radiation, motion artifacts, and the interaction of different tissues. Hence, pre-processing is needed to filter out the unwanted noises in EMG.

Eye Tracking

Gaze, defined as the direction to which the eyes are pointing in space, is a strong indicator of attention, and it has been studied extensively since 1879 in psychology, and more recently in neuroscience and in computing applications for recording and studying human visual behaviour [DUCH2002]. Many studies have used eye-tracking for usability evaluation of the interface [JACO2003, EHMK2007]. Goldberg and Kotval [GOLD1999] were among the pioneers of investigating the usage of eye tracking measures when browsing different types of web-pages. The main measurements of eye tracking used during HCI evaluations are “fixation times” (moments when the eyes are relatively stationary, taking in or “encoding” information), “saccades”, which are quick eye movements occurring between fixations, “scanpath” (describes a complete saccade-fixate-saccade sequence), pupil size and blink rate. Goldberg et al. found that a higher number of saccades is an indicator for a poorer interface [GOLD1999], and in their overview on eye-tracking research in HCI and usability. Jacob et al. stated that the mean fixation duration can be an indicator of a participant’s difficulty in extracting information from a display [JACO2003]. Eye blink rates and pupil size yield meaningful information about task demands and level of fatigue and can be used as an index of cognitive workload [ALLA2004]. Studies on eye-blink rate have shown that it is inversely correlated with attention or mental load, i.e. the lower the blink rate, the higher the mental load or attention [GOLD1999, BRUN2002, BROO1996]. Larger pupils may also indicate more cognitive effort [MARS2000; POMP2003]. Sharma et al [SHAR2014] gave an overview of the usage of eye tracking in other non-medical areas. They concluded that the future research should look for the correlations of usability problems that have to be related with the specific patterns.

Kurzhal et al. [KURZ2015] discovered that one of the main advantages of the eye tracking is that the eye movement data contains the information about where and when participants waste time when performing a given task. Those “wasted time” can be a hint(s) of possible design issues. Besides, researchers are able to construct a model of the cognitive architecture regarding the user's behaviour using the eye tracking measures [KURZ2015]. This aspect is not contained in other traditional measures, e.g., video analysis or using the log files.

Sitting posture is another important measurement in the evaluation of input devices as a proper sitting position may lead to long-term comforts. For instance, in evaluating various touch interfaces, Bachynskyi [BACH2015] reported the similarities and differences of sitting postures using 1) a motion capture system and 2) a custom-built chair which measured the forced applied by the user in various directions while using different input devices.

2.5 Conclusions after literature review

Section 2.1 of this review summarizes the roles of users in a (semi-automatic) segmentation process. It was identified that user interactions were required during three main phases of segmentation: 1) the initialization; 2) the post-processing corrections and 3) the intermediate segmentation corrections. In our study, we want to address that in practices, the "fully automatic segmentation" do not exist as all the segmentation workflows requires the involvements of users, at least in checking the outcomes of the computational algorithm and making necessary corrections.

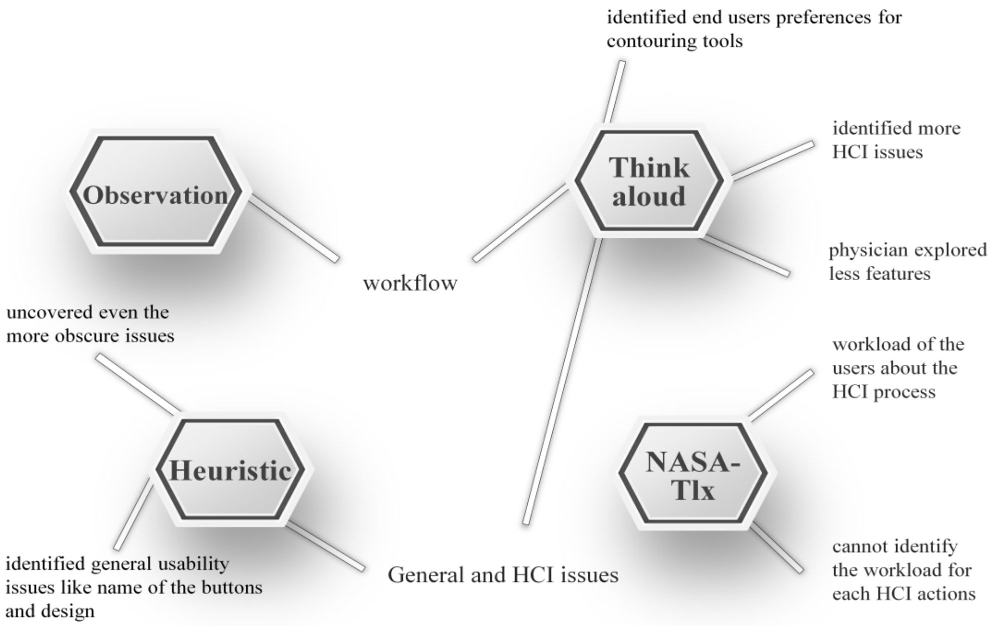
Section 2.2 reviews the types of user input tools that support the user interactions. These are classified as four different categories such as point based, line based, area based and volume based tools. From the literature, it can be identified that even though many efforts were paid to develop different types of user input tools [LI2012, WANG2012], there is no clear recommendations about what are their influences on the process and results in the context of medical image segmentation. Also, the knowledge about the effects of various types of user interactions on the results and the HCI process during segmentation is limited. Hence future studies should be conducted in order to explore the effects of different user input tools on the segmentation process and the results.

The last part of Section 2.2 reviewed various input devices. A number of input devices are available within radiology and radiotherapy. Other than the regular pen and the mouse, tablet-PC, multi-touch and gesture-based interactions could be beneficial in radiotherapy therapy treatment planning. Hence there is a need to investigate its potential. In the field of radiotherapy and radiology, few research had focused on the HCI process of using these input devices. Thus future study on investigating the impact of user input devices on the HCI process and the segmentation results can be beneficial.

Section 2.3 gives an overview of the most common evaluation methods that are used in UE and HCI. Literature survey indicates that there are many evaluation measures that have been identified and used in UE and HCI for various purposes. But there is no clear protocol on which evaluation method should be used during which stage of the design process. Besides, the questions of what are the inter-relations between these measures in using the input tools and devices, and how to combine the outcomes of those measures to make a proper selection of HCI tools and devices and to identify design issues in their usage remain to be answered.

Field Research

3



This chapter is based on:

Anjana Ramkumar, Yu Song, Wiro J. Niessen, Pieter Jan Stappers. "Design issues of the existing medical image segmentation software: A case study using observational, heuristic and think-aloud methods". Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare June, 2016, 5(1): 1-8.

Anjana Ramkumar, Yu Song, Edit Varga, Wiro J Niessen, Anne Laprie, Ben Rowland, Adinda Freudenthal. "Comparison of Heuristic Evaluation and Think Aloud Methods A Study in Radiotherapy Contouring Software". Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare 2014 June; 3(1): 230-237.

The previous chapter showed that a large number of HCI approaches, HCI tools and HCI input devices are involved during segmentation process. In this chapter, we used the observational research, the heuristic evaluation method, the think aloud method and the NASA-TLX questionnaires to: 1) analyse the workflow of radiotherapy segmentation systems; 2) discover possible usability and HCI design issues of current segmentation systems in order to identify the requirements for future interface design; and 3) explore the abilities and limitations of various evaluation methods. This chapter is arranged as follows: Section 3.1 illustrates the methods and materials that are used for this study; In Section 3.2, the workflow analysis based on the observational method, the evaluation results of the heuristic evaluation method, the think aloud method and NASA-TLX questionnaires are presented; In Section 3.3, the evaluation results of the workflow and the interfaces, as well as the evaluation methodology are discussed. Section 3.4 concludes the chapter.

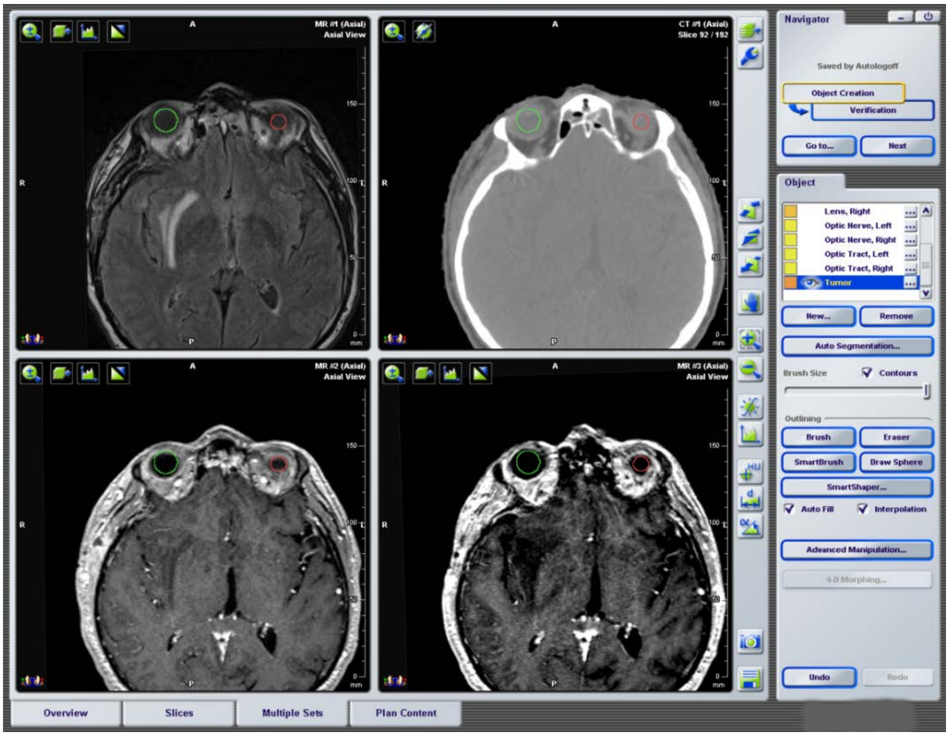
3.1 Materials and Methods

Evaluators

This study was conducted at the Department of Radiation Oncology, University medical centre Freiburg, Germany and Institute Claudius Regaud (ICR), Toulouse, France. During the observational study 8 physicians were observed. For the think aloud method, five physicians with 3-7 years of experience in radiotherapy segmentation joined the evaluation. In the heuristic evaluation, due to the needs of in-depth medical knowledge, instead of human factors (HF) experts, two evaluators, both with HF knowledge and strong background of radiotherapy, conducted the experiment. To bridge possible gaps, both evaluators evaluated the system, and consulted each other before any final decisions were made. If a consensus could not be reached, local experts were consulted for a conclusion.

Software Systems

Three commercial software systems that were being used for radiotherapy segmentation in clinical practices had been considered for this study. All the physicians were familiar with all the three systems. Figure 3.1(1) shows the screenshot of the “Object creation” window of System. It consists of 5 different tabs namely: overview, slices, multiple sets and plan content. Among them, only the slices and multiple sets are used for contouring. In the figure, the tab multiple sets was active and divided divided into four quadrants. Each quadrant shows different



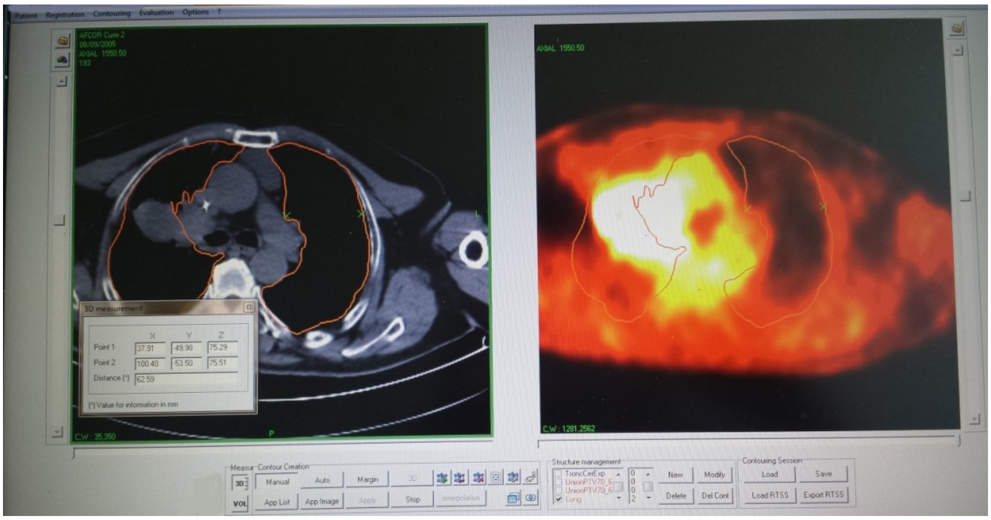
1) System- A: multiple set window of object creation interface is presented

image modalities, for example on top left was the MR image, top right was the CT, bottom left was the MR T1 image and the bottom right was the MR T2 image. Depending on the institution, the selection of the image modalities may differ. A tool bar lied at the left of the windows for scrolling the images up or down, zooming of the images, panning of the images, adjusting the intensity of the images, measuring of Hounsfield Unit(HU), angle and distance. The right side of the interface shows the OAR template list, auto segmentation button, and below are the tools that were used for manual contouring. At the bottom functions such as screenshot, save and undo are presented.

Figure 3.1(2) shows the screenshot of the System B. On the top row (below the yellow colour bar) of the interface is the toolbar. Some of the commonly used tools are nudge tool (named as pearl tool), free hand tool etc. The centre of the interface shows a CT image. On the right-hand side of it are the tools for adjusting the contrast of the image. Below the image is the organ list manually created by the physician before segmenting any organ.

Figure 3.1(3) shows the interactive contouring interface of system C with two

2) System- B



3) System- C

Figure 3.1: Interface of the three software systems

images. The windows on the left side shows the CT image and other shows the PET image. Below the image there are three sections. The first section (on the extreme left) shows the "manual contour creation" tab and it contains the toolbox for creating the contour. The centre section is the "structure management" which contains the ROI list and the possibility to add or delete and existing organ on the list. The third section is the "contouring session" which has the option to load or export a structure set.

The systems A and B were being used every day by physicians for segmenting OARs. The system C is used only for the datasets that are considered for clinical trials.

Datasets
A lung and brain dataset is used for segmentation in all the three systems. The usage of all the datasets were approved by the ethical board of the University medical Centre Freiburg.

Evaluation Method
Four types of evaluation methods were used. Before starting the whole study, the physicians signed the consent form.

Observational Method: As an initial step, one researcher did two weeks of observational study during which the users were observed in their clinical settings while performing the segmentation task. The observational study was conducted to understand the workflow of the segmentation systems. With this method, we also identified that system A, B, and C have some design and usability issues. The duration of the whole observational study for the two cases took about 1.5-2 hours for each physician.

Heuristic Evaluation (Expert): The evaluation was done with a heuristic evaluation checklist developed by Zhang et al. [ZHAN2003] (Table 3.1), which was adapted from Nielsen [NIEL1994]. Using these heuristics, each interface that physicians may encounter in the segmentation task was assessed. The duration of the whole heuristic evaluation was about 6 hours for each software, which is much longer than a "normal" contouring task since: 1) The evaluators went through all possible interfaces, not all of them are needed in conventional contouring tasks; 2) The time spent for consultation with medical experts was also included.

Table 3.1: Zhang's heuristic evaluation checklist

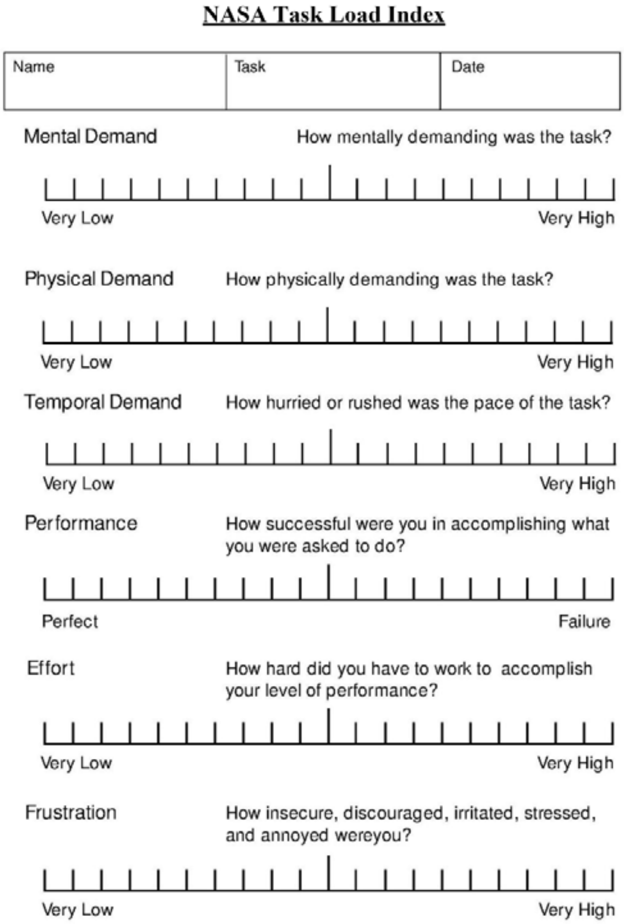
Heuristic	Explanation
Consistency	Users should not have to wonder whether different words, situations, or actions mean the same thing.
Visibility	Users should be informed about what is going on with the system.
Match	The image of the system perceived by users should match the model the users have about the system.
Minimalist	Any extraneous information is a distraction and a slow-down
Memory	Users should not be required to memorize a lot of information to carry out tasks.
Feedback	Users should be given prompt and informative feedback about their actions.
Flexibility	Give users the flexibility of creating customization and shortcuts to accelerate their performance.
Message	The messages should be informative enough such that users can understand, learn and recover from errors.
Error	Design interfaces that prevent errors.
Closure	Users should be clearly notified about the completion of a task.
Undo	Users should be allowed to recover from errors.
Language	The language should be always presented in a form understandable by the intended users.
Control	Do not give users that impression that they are controlled by the systems.
Document	Always provide help when needed

Think aloud : In the evaluation process of using think aloud method, users were asked to express their thoughts while doing the segmentation task. Semi-structured interviews were conducted at the end of the study. Video-recording was used for detailed analysis of the study. For each case, the duration of the study was about 1 hour. The users were asked to segment the OAR on the datasets using the systems A, B and C. There were no strict protocols in doing the segmentation as they were asked to do in the similar manner as they do in their daily clinical practice. The physicians were given the freedom to use the automatic or semi-automatic segmentation algorithm as they used in their normal routine.

NASA-TLX: NASA-TLX [HART1988] questionnaire (Figure 3.2) was used as a subjective measure to determine the workload of the user during the segmentation process in using different systems. Physicians were asked to fill the questionnaire each time when they finished a task in a system. The questionnaire was divided into 21 subscale dimensions, each ranging from low to high workload. The scores derived from the 21 subscale are then averaged to produce an overall workload index.

Data Analysis

The video recordings were manually transcribed into transcripts as Microsoft word files. Coding categories were assigned to usability issues identified in the transcripts.



NASA Task Load Index

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Mental Demand : How much mental activity was required (deciding, thinking, calculating, remembering, looking, searching etc.)?

Physicial Demand : How much physicial activity was required (pushing, pulling, turning, controlling, activating etc.)?

Temporal Demand : How much time pressure did you feel during the task?

Performance : How successful you think you were in completing the goals?

Effort : How hard did you have to work (physically and mentally)?

Frustration : How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed, complacent did you feel during the task?

Figure 3.2 NASA-TLX questionnaire [HART1988]

3.2 Results

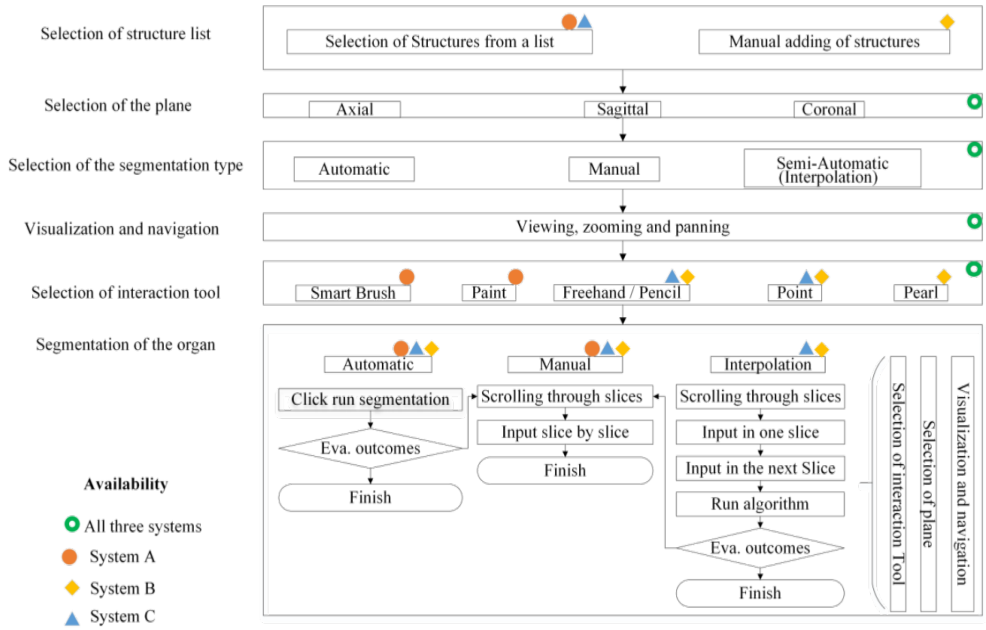


Figure 3.3: Workflow and main features of different segmentation systems

Workflow: Using the observational method and think aloud method, we observed the segmentation task done by users in different commercial software systems and identified the workflow of those software systems. Figure 3.3 shows the workflow and the components involved in those systems. The main tasks, which are indicated on the left of the workflow was same for all the systems. At first, users were asked to choose the structures that they wanted to segment. After selecting the structures, users can choose the plane in which they wanted to segment. If required, the window levels, and the zoom level of the images were adjusted after selecting the plane. Depending on the structure, users selected among the following segmentation methods: The first method is the automatic segmentation where user's task was to click the "run segmentation" button. This method is considered to have no/few user interactions. Each automatic segmentation has different algorithm(s) associated with it; The second method is the semi-automatic segmentation method. The semi-automatic segmentation methods available in these systems was called as "Interpolation". Using this method, users were required to segment manually in two different slices and the slices in between were segmented automatically. However, in all three systems, both the automatic and semiautomatic methods required a lot of manual post processing. The third method is the manual segmentation. In manual segmentation, users were asked to segment manually in each slice using the available segmentation tools. Different systems had different HCI tools for performing the segmentation.

Figure 3.3 shows some of the tools that were commonly used by users during the study. Among those tools, the smart brush tool works based on the threshold of the region. The paint brush, the free hand drawing and the pencil tools were different types of "free hand" drawing tools. The only difference between the paint brush and the other two was that the size of the brush was adjustable. Using the point based tools, users needed to mark the points on the border of the organ and the points were connected by lines according to the sequences of the “clicks”. The pearl tool is a type of nudge tools. Using this tool, it was able to expand the contour when the tool was inside the contoured region and shrink the contour size when it was outside. Selection of the tools, navigation through the dataset and plane selection was required throughout the segmentation process as shown at the right of Fig.3.3. The colours indicate the availability of the features in different systems. If a particular feature is labelled by green, it was available in all the three systems. The outcomes from the heuristic method and think aloud were categorized based on the steps identified from workflow analysis. It is worth mentioning that in the use of the heuristic method, physicians were not involved directly, however, their feedback is addressed in the use of the think aloud method.

1. Structure List

Heuristic method

System A: *Consistency*- Region Of Interest (ROI) had a long list of objects even after selecting the type of treatment.

Flexibility- No option to remove multiple objects at the same time.

Minimalist- Extra template information is unnecessary. To create a new object, users must provide a name and select a type of structure from a limited list. Structures on the list are objects that are stored in the anatomical atlases. However, obvious mismatches could happen between the structure selected and the name provided that it becomes invisible from the main window.

Visibility- Structure type is invisible in the main window.

Error- Those mismatches may lead to error in future treatment planning.

Message- The user should be alerted if the chosen name and the corresponding structure type are different.

System B: *Memory*- There was a pre-defined list which was not complete. Hence the users need to manually enter almost every time and need to remember which organs needed to be included for different tumour type.

System C: *Flexibility*- The complete structure list was not visible to the user in one window. They had to scroll up and down the list every time to search for the structure. The users cannot see the complete list without scrolling.

Think aloud

System A: Region Of Interest (ROI) had a long list of objects even after selecting the type of treatment. After selecting the cranial treatment option, users expected to see only the structures that were in the cranial part of the body, but the system provided a long list of structures that were outside the cranial region.

System B: The users found it a bit time consuming as they had to enter many structures manually which involved more user interactions.

System C: The users did not mention anything regarding the structure list.

2. Visualization

Heuristic method

System A, B, and C: *Flexibility*- The system should have the flexibility to include images as many as possible e.g: the brain segmentation might requires up to 8 imaging modalities. However, viewing of all the images simultaneously was not possible.

Think aloud

System A: It was not possible to view different image modalities overlaid on top of each other. It was only possible to view them side by side in different windows. Users had to frequently switch among different images for drawing a satisfied contour. The switching tasks were considered as time consuming.

System B and C: It was not possible to view more than 2 imaging modalities at the same time.

3. Tools and Buttons

Heuristic method

System A: The smart shaper tool had 2 options as “deform” and “move”. If the “move” option was selected, all the contours on that particular slice was moved. However, if the “deform” option was selected, only one contour got deformed on which the cursor was being placed.

Match The user expected that either the “move” option should move only one of the contour or the deformation option should deform all the contours in the particular slice. There was no option to delete the contour by one mouse click / keystroke.

Flexibility – There is no shortcut to delete the contour.

Consistency- In two tabs (Overview and Slices), the advanced windowing button is located only at the right of the tool bar. However, in another Tab (multiple set Tab), this button appears in each quadrant at the top left corner.

Minimalist- unnecessary, distracting extra buttons.

Measuring distance and measure angles

Consistency- In measuring distance and measure angles, the user only can draw 3 lines in one slice. However, they expected to draw more lines to specify angles (limited by the system).

Match- Measuring distance option did not work according to user expectations. The measured distance function was achieved by clicking on two different positions. But users cannot move existing measurements. If the option was switched off, those lines were deleted automatically.

Memory- User had to remember the distance measures as it did not appear after switching to other functions.

Undo- If the user drew a line and switched on and off the same option, he/she did not get the line back.

Match– Mismatch between the design and the functionality. Full screen icon looks like a zoom-in/zoom-out tool rather than an option to maximise the screen.

Consistency– Similar symbols are used for both the zoom-in and zoom-out functions.

Match-Users’ expectation was not fulfilled. No consistency between the design of windowing button.

Consistency- Basic and advanced options should have similar style.

Minimalist- Unnecessary, distraction of extra buttons. Double positioning of the plane selection button on multiple set tab. In the rest of the tabs this button appeared only at the right toolbar. However, it also appeared on each quadrant at the top left corner. The name appears at different locations.

Undo- The users were not able to recover from their errors.

Match- The names did not match to the expectation of the user. On the “new object selection” window the user got two options as the “single object” and “multiple objects”. However, the multiple objects is a template so it needs to be replaced as “multiple objects template” or “template of multiple objects”.

Minimalist - No short cut for adjusting the grey scale value. Users had to perform a lot of mouse clicks to adjust the grey scale.

Match - Panning option was not flexible. The user needs to zoom a bit to pan, otherwise panning was not possible.

System B: *Consistency* and *Minimalist*: The users cannot draw two different contours using the pear tool. They had to extend the first contour and then delete the connection in-between. This was an unnecessary extra work for the users.

Match - Automatic contouring symbol did not visually represent as “automatic contouring”.

Match - ROI magnification symbol is not clearly visible. It did not match the purpose.

Match - Symbol of the “Help contour” did not match the purpose.

System C: *Consistency* and *Match*-Naming of the buttons were similar and hence there was more confusions: for instance, “Delete contour” – deletion of all the contours in a structure, “Del contour”- Deletion of one single point in a contour. “Delete”- Deletion of a contour in single slice.

Closure - Users had to struggle a bit to close the contour while using the pencil tool.

Match - 3D button– It was actually a scale but the naming “3D” did not match the scale. It was named as 3D because it showed measurements in all the x, y, z direction.

Think aloud

System A: Physicians mentioned that the smart shaper tool was not useful in segmentation because it worked according to the intensity due to which sometimes it contoured some unnecessary organs as well as shown in Fig.3.4a.

Undo function was missing. The user wanted to undo the performed actions but he/she was not able to do it.

System B: According to the users there were many tools on the interface. However users used only one or two tools and the most commonly used ones are the *free hand drawing* and *pearl tool*. Due to the large number of tools, the interface was little bit squeezed and thus it was difficult for users to search most of the items.

Physicians 1 mentioned that “Delete button looks like water closet….. hahahaha… ”. Physicians was asked to define the bolus function and they replied “I do not know what it is”.

All the tools symbols were (or at least looks) small and most of it needed to be clearly defined.

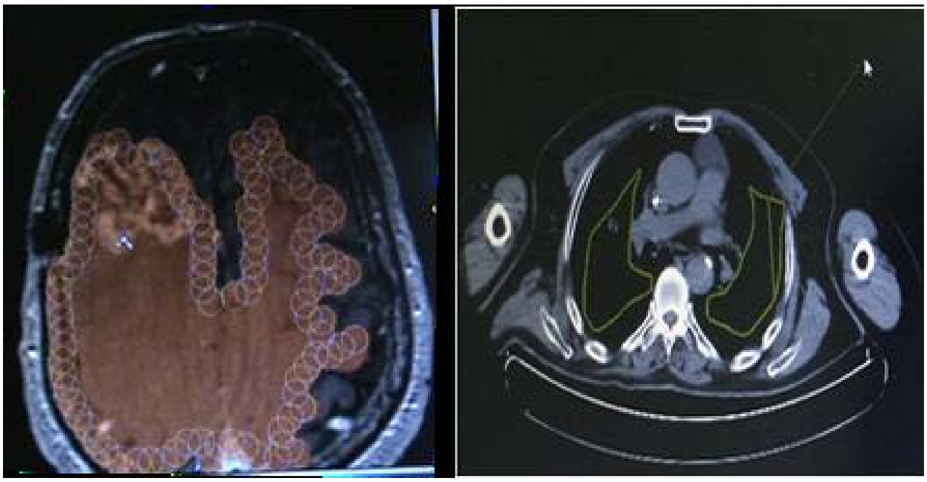


Figure 3.4: Usability issues with the tools. The left image (a) shows the issues with the smart shaper tool. The right image (Fig 3.4b) shows the problems with closing of the contour.

System C: Many buttons were named similarly. Even the users where confused with the similar naming of the buttons and they were exploring what each button did. Modification between two points using *pencil tool*- As shown in Fig. 3.4b, users had to struggle a bit to close the contour.

4. Segmentation of the organ

Heuristic method

System A: <u>*Consistency*</u>- Inconsistency with user's expectations. (When one of OAR was selected and the automatic segmentations was applied, user expected that only that one organ was segmented automatically. However, when user performed automatic segmentation function, all "objects" were segmented automatically).

<u>*Consistency*</u>- Sometimes the system interpolated between 2 slices and sometimes 4 slices. There was no clear indication and no consistence in the algorithm.

System B: <u>*Flexibility*</u>- If user contoured in the first and the seventh slice, the interpolated contours in-between was displayed as dotted lines around that structure. Only after converting the dotted lines to contours, modifications were possible. If two adjacent slices were wrong, it was not possible to correct all of them together. Hence there was lot of user interaction and mouse clicks involved in each slice.

System C: <u>*Consistency*</u>- When automatic segmentation method was used, the outcome was good for certain structure like spinal cord. However, for many organs, post- processing tools were required in order to correct the outcome.

Think aloud

System A: Automatic segmentation works on all the structures just by clicking the automatic segmentation function. The user expected this function can be applied only on a particular OAR. However, there was no possibility.

System A, B and C: Interpolation was not optimal because of required rich HCIs.

5.Others

Think aloud

Use of mouse and pen-tablet at the same time for contouring and observing results, which is physically challenging but currently hysicians do not have any other options to contour.

In the heuristic evaluations, two evaluators discovered a total of 41 violations together with all the three systems. 26 violations were identified in system A and seven violations were identified in system B and C, respectively.

NASA TLX: Users' workload in using all the systems is shown in Fig.3.5. It can be seen that the workload level of system C is higher in almost all the categories. The mental demand was higher for system A and the physical demand was almost the same for system B and C.

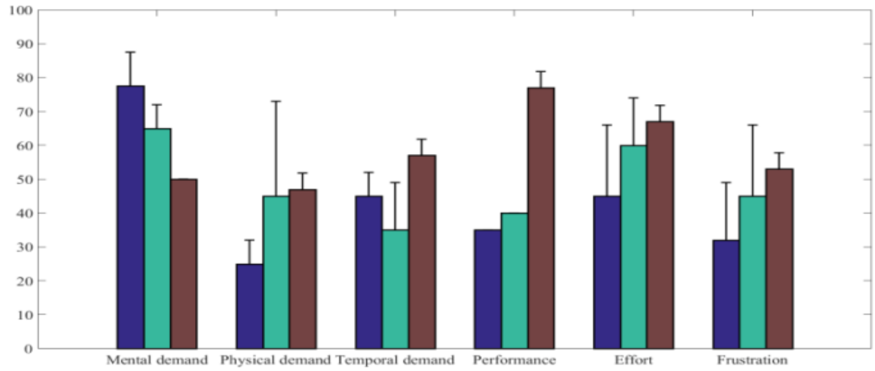


Figure 3.5: NASA-TLX workload of all the user in using all the systems. Violet: system A, Blue: system B and Brown: system C

3.3 Discussions

In this study, we used the observational, heuristic, think aloud and NASA-TLX methods to understand the segmentation workflow, to identify the usability and HCI issues in existing commercial software systems, and to understand the workload of the users in using three systems. Using a single method may not be enough for comprehensive usability studies [KOUT2007]. The observational method was very useful during the workflow analysis and identified some small usability issues. The adapted heuristic evaluation method uncovered even the more obscure issues with the interface. In the protocol of heuristic evaluation, both evaluators discussed between themselves and evaluated each possible violation. Only after coming to an agreement, this possible violation was marked as a violation. Many studies have been conducted by using only the think aloud method and later the heuristic checklist to identifying the issues [ROGE2013]. Hence, the role of human factor experts becomes less important in those studies. Similarly, heuristic approach is often combined with observational studies. The main advantage of our technique was that conclusions made regarding each violation were consolidated in the discussion instead of one person’s decision. The outcomes of the heuristic evaluation methods proved that: 1) using our setup it was possible to explore and comment on each and every functionality of the interface such as button designs, size, etc., 2) most of the user interface violations identified by think aloud method had been confirmed by the heuristic method. This clearly proves that the evaluators with HF and professional knowledge could identify more problems than the ones without professional knowledge regarding the particular type of software.

With the think aloud method, users were able to compare their system to other contouring systems. Also, it was easier to identify many HCI issues on the contouring tools, such as the brush, the pearl tool, etc., and on the input devices. Most of the results were based on the experienced end users' daily practice, which showed fixation on specific strategies. Hence, they did not explore less unknown features, even when they were asked to do so. According to Yen at al. [YEN2009], if a usability problem does not impact end user's task completion, it may be less influential over time as they get used to the system. With the NASA-TLX, it was identified that except the mental demand, the rest of the workloads were higher using system C. However, the heuristic method identified more issues on the system A. This could be due to the fact that the users frequently use system A and B rather than C. Hence, the slow learning curve might have led to higher workload level.

With the four methods, this study identified several usability and HCI issues using the three-different software interface and some requirements have been proposed as:

General Usability issues

Structure list: All the systems have a list but should be customizable depending on the context.

Visualization: During some OAR segmentation tasks, physicians visualize different image modalities such as pre-operative images, post-operative images, etc. Currently, physicians can view only two images side by side and hence they continuous opening and closing of one image and then repeating the same for other images, which is physically and cognitively demanding. The future design should include different visualization techniques with which the users can visualize multiple image data at the same time.

Other Design issues: The naming of the button functions must be clear so that physicians will not be confused. The general interface issues are very basic design issues and these could have been avoided if proper usability testing was done at the early stage of the design process. However, at least the basic design should be consistent.

HCI Issues

***HCI tools*:** Unnecessary and unwanted regions were segmented by using the smart brush tool (Fig 3.4a). With majority of the tools e.g. free hand, paint, pencil, etc., the user must input along the boundary of the organ precisely and there was no other efficient technique available to give the user input. Using point based tool was very time consuming compared to other drawing tools. Also physicians preferred to have opposite functions on the same tool e.g. paint and erase so they do not have to switch between many tools. This might also reduce the number of user interactions. Hence, the future design should involve a tool through which the user can give their input other than the organ boundary only and which could incorporate the opposite functions in a same tool (e.g. paint and wipe).

***Input device*:** It was observed that physicians used multiple devices for performing functions like scrolling, contouring etc. The future design needs to synchronize the interface and the input device, so that users can use only single device instead of multiple devices at the same time.

Segmentation

Automatic segmentation: With the automatic segmentation, users considered it as time consuming due to imperfection with the algorithms and a lot of post-processing corrections were needed. Also the current post-processing is done slice by slice manually, which was time consuming. Hence, the future design should include necessary post-processing correction tools which could perform efficient segmentation corrections.

***Semi-automatic segmentation*:** Similar to automatic segmentation, interpolation also required many user corrections. Systems A and C were not consistent in giving the outcome. Some physicians found it hard to understand the interpolation process and hence they did not use the interpolation function of this system. However, the system B showed real time feedback and users considered it as advantageous. Hence, the future interpolation design must give consistent real time feedback. Also with all the systems apart from the contour input method, there was no other way to give their input. The smart brush tool was the easiest tool, but it did not segment the whole organ. Hence an efficient way to give the user input is required.

Limitations

In the design of this study, we also identified some limitations. The physicians involved with the think aloud study were not native English speakers, language barrier may also prevent precise descriptions of the problems.

3.4 Conclusions

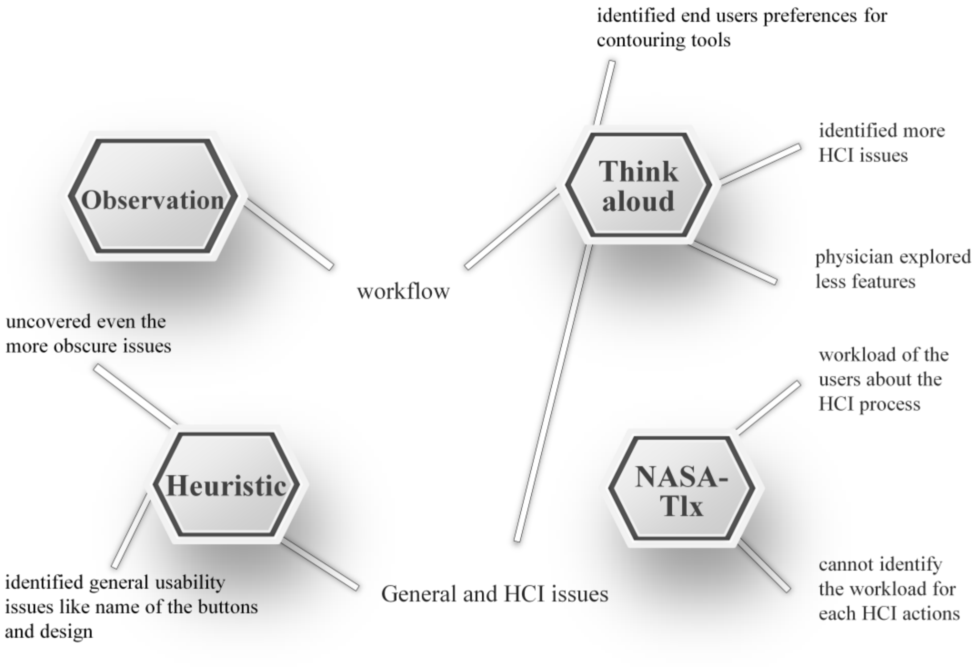


Figure 3.6: Overview of various evaluation methods and their benefits and disadvantages

In this chapter, we analysed the existing radiotherapy planning software interfaces using four different methods. The observational and the think aloud methods were very useful in identifying the workflow of different systems. The heuristic evaluation method uncovered more specific issues with the interface design. An added advantage of this method is that as both the evaluators have background and knowledge in radiotherapy they were able to explore and comment on each and every functionality. The think aloud method identified more HCI issues compared to general usability issues. Based on the comparison of the outcomes of, it is suggested that in the process of improving usability of a radiotherapy contouring interface, the think aloud method can be applied to explore the HCI issues. The heuristic evaluation can be applied in designing the details of the interface such as size and design of the buttons, design of the windows, etc. In addition, it is also recommended that in heuristic evaluation, using multiple evaluators with both professional and HF knowledge may identify more usability problems. This study identified several usability and HCI issues and proposed seven main requirements regarding future radiotherapy planning software:

1) System must have a predefined structure list and have the possibility to customize it;
2) Visualization of multiple image data should be possible at the same time;
3) Naming of the buttons should be clear and consistent;
4) Automatic segmentation must include necessary and efficient post-processing correction tools;
5) Semi-automatic segmentation must be consistent and needs to give real time feedback;
6) HCI tools needs to have opposite functions on a same tool;
7) Input device design must synchronize well with the interface design.

Prioritization and Design Focus

4

Part of this chapter is based on

Anjana Ramkumar, Edit Varga, Anne Laprie, Wiro J Niessen, Adinda Freudenthal

A pilot study of Pen-Tablet interaction in radiotherapy contouring using orthogonal and non-orthogonal views.

Design of Medical Devices Conf – Europe (DMD'13 – Delft), October 2013

In previous chapters, based on literature research and by studying current commercial segmentation software solutions, a considerable number of design issues, possibilities and wishes were gathered about the interactive segmentation workflows, HCI tools, the input devices, etc. In this chapter, the design issues, possibilities and wishes collected are prioritized for setting up the design focus. As mentioned by Freudenthal et al. [FREU2008], improving one part first and gradually expanding the improvements allows medical professionals to steer the system in the desired direction. Prioritization could prevent from just building a fancy technological solution which is incomplete, or has functions which are superfluous. In this Chapter, Section 4.1 presents the prioritized design requirements, Section 4.2 presents a pilot study, which was used to consolidate the requirements and in Section 4.3, the design focus was set.

4.1 Prioritization of design requirements

Chapter 2 summarized the roles of users in the interactive segmentation process. User inputs are often needed during three main phases of the segmentation process: 1) for initializing the computational algorithm; 2) for post-processing corrections and 3) for correcting intermediate segmentation results. Depending on the design of an interactive segmentation method, user input approaches, user input tools and input devices play important roles in (part of) these three phases.

Input approach

Two types of user input approaches are often used: the direct and the indirect approach, depending on if the user is asked to directly specify the output of the interactive segmentation method. The direct approach is the most popular approach in current segmentation software solutions, however, it could be physically demanding for the larger organs such as the lung. Regarding an indirect approach which user needs to roughly specify the locations of the organs, the effectiveness and efficiency this approach are yet to be explored. It is worth mentioning that the effectiveness and efficiency of either the direct or the indirect approach strongly depend on the incorporated segmentation algorithms. As this thesis does not focus on the computational algorithms, a combination of graph-cut and watershed-based algorithms which was able to facilitate both approaches [DOLZ2014a, DOLZ2014b], was adopted to test different approaches.

Input tools

There are many different types of user input tools used for radiotherapy segmentation as summarized in Chapter 2. Using various tools users can give their inputs through different options, e.g., the menu option selection, the pictorial input on an image grid, and the parameter tuning. The menu option selection and the parameter tuning inputs are strongly associated with the algorithms that are used in the segmentation methods, and the effectiveness and efficiency of user inputs mainly depend on his/her understanding of the algorithms. The pictorial inputs are different where the user can play a major role by directly feeding the algorithms with drawn points/contours/areas based on their medical knowledge. Four types of pictorial inputs were identified in Chapter 2: the point based, the line based, the area based and the volume based inputs. Among them, the line based input is the most commonly used tool in radiotherapy segmentation, few literatures discussed other inputs in medical image segmentation, especially in the segmentation tasks for radiotherapy.

Input devices

In Chapter 2 and 3, besides many user interface design issues which should be addressed in the interface design, we also identified that the input devices must be synchronized well with the interface design. It was clear that mouse and pen on pad are the most common input devices used. However, in many other fields, pen on screen and touch screen were used as effective input devices. During field research, physicians also mentioned that they would like to try new input devices for image segmentation.

The objective of this research presented in this thesis is to find the effective and efficient HCI for interactive image segmentation. By summarizing the findings in the literature study and field research, we prioritize the requirements as investigating the impact of: 1) different input approaches; 2) different types of pictorial inputs tools and 3) different input devices used in the segmentation process and the results. However, two aspects should to be further confirmed before the we set the design focus: 1) what type of information should be displayed on the interface for segmentation and 2) how many input devices shall we study. In the following section, a pilot study was conducted to find answers to the above questions.

4.2 Pilot study

Physicians used to contour organs-at-risk in orthogonal views of the dataset, especially in the axial view. Currently they mainly use the mouse to finish the task, sometimes using pen on pad was also observed. In this section, a pilot study is conducted to investigate: 1) the possibility of using other views in the interface and 2) the possibility of using different HCI devices in the study. As a pilot, we did not go for a full design. Instead, we designed and implemented a vertical prototype with the functions that were needed for the test. Regarding these functions, the prototypes were considered as high–fidelity prototypes [RUDD1996]. The evaluation method was also primitive, mainly based on subjective measures such as observations, semi-structured interviews, etc. However, we expected such a pilot is able to deliver useful information about the potentials of different types of design. Those potentials will help setting the design focus, and it will be further investigated in the full-scale study.

4.2.1 Materials

Two vertical prototypes were developed during the pilot study. Prototype 1 utilized a Wacom Cintiq 21UX screen as a pen on screen input device. The screen was connected to a laptop which ran the software. Figure 4.1 shows the pen on screen input device. The interface had the option to contour the images on either orthogonal planes (axial, sagittal, and coronal). Also, there was an option to visualize the 3D volume that was generated based on the contours, or alternatively fill in the 2D contour in the 3D view to enable a better interpretation of the 3D anatomy (Fig.4.1).

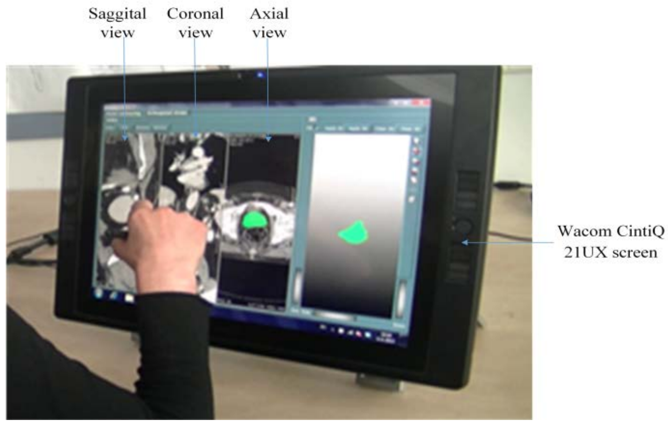


Figure 4.1: The prototype 1: Pen on screen input device

Prototype 2

Prototype 2 was installed on a Windows 7 tablet of 10 inches and was connected to a 21.5 inch monitor screen. Besides, a pen input, the tablet also had a touch screen where the physician could use their fingers for contouring. Both the tablet screen and PC showed three sub-windows; the first visualized the plane which the physician selects; the second one was for viewing the 3D image volume and the third one visualize the 3D contoured volume.

Figure 4.2: Touch screen input device

The main purpose of both the prototype design was to understand if pen on screen and finger input can be used as an input device for the future studies. Also we had designed the orthogonal view in prototype 1 and non-orthogonal views in prototype 2, to understand if non-orthogonal view has some additional benefit than using the orthogonal view and also to find out the type of information that should be displayed on the interface for future prototype design.

Both the prototypes were developed in Mevislab 2.2.1 [MEVI2016]. Prototype 2 used NDI Aurora technology for tracking [AURO2016]. A sensor was attached to the tablet, which enables the physician to select different planes on the tablet by rotating it. Upon finding a desired plane, the physician will be able to stop the navigation mode, lay down the tablet on the table for contouring. The tablet was connected to the computer screen such that the images on the tablet appears on the screen, for better visualization of the images. Two different anonymised patient datasets were used for both the prototypes. Utilization of the datasets for this study was approved by the Ethics committee of Institut Claudius Regaud, Toulouse. Prototype 1 had whole CT data from the brain till feet. The other dataset was restricted only to the abdominal region.

Participants

Two annotators participated in this study. The first one was a radiation oncologist from the Institut Claudius Regaud, Toulouse, France, with 15 years of experience in contouring CT images. Participant 2 was a medical physicist at the Radiology department of Erasmus medical centre, Rotterdam, The Netherlands, who was in her final year of medicine education and doing research in liver tumour contouring on CT images. She had only few months of experience in contouring, however, in these months she had intensively training, i.e., contoured hundreds of datasets. Both physicians had experience in contouring only on the orthogonal views, mostly in the axial plane.

4.2.2. Method

Experimental task

Task1 was to manually contour the prostate using Prototype 1. The participants were given the freedom to choose any plane as per their wish. The second task was to segment the bladder using Prototype 2. The reason for selecting the bladder was that it is one of the organs which could be easily contoured on any views because of its spherical shape. They were explained how to use the prototype and instructed to contour on any of the oblique planes.

Test setup & protocol

The test was conducted in a large meeting room, where the prototypes were placed in the opposite ends of the room. There were two test leaders assigned to the two participants. Thinking aloud, semi-structured interviews and guided expert discussion were used as evaluation methods. Video-recording was used for detailed analysis of the study. Before performing the tasks, participants were given a very short demonstration of the two prototypes in which their functionality was explained and questions were addressed. Each task took 10 minutes on average and the whole study with breaks and end discussion took approximately an hour. As the task was done during a workshop, the time was limited to 10 minutes.

4.2.3 Results

1) Contouring

Prototype 1

Very confident and easy contouring in all 3 orthogonal planes. Participant 1 expected interpolation of contours on the next slice which was not provided. Contour filling was much appreciated.

Prototype 2

Easy contouring in standard view. Even though it was difficult to understand the anatomy in non-axial views, both physicians managed to contour 2 slices. Participant 1 expected to select a single contouring plane direction and to contour that plane, which was not an option in our prototype.

2) Interactions

Prototype 1

Prototype 2

Use of pen on screen was very intuitive. Physicians preferred the 2D and 3D images with contours, on the same screen. Filling of contour volumes was seen only in axial plane, but physicians expected it to appear in all the 3 orthogonal images.	The concept of using touch screen was much appreciated by the Physician 1. Physician 1 preferred to use her finger for contouring over using a pen. The contour filling option was not used by both physicians because they did not reach to that point as the task time was restricted to 10mins.

4.2.4 Discussions

In this section we discuss the outcome from the two designed prototype. First, we discuss the information that needs to be presented on the UI and then we will discuss about the input devices that can be used for future testing. Finally, reflections of this study are discussed.

1. **Information be presented on the UI**

Interpretation of the anatomy

Normally, physicians use orthogonal slices of CT datasets to check the tumour and to see which slices need to be included for contouring. The use of non-orthogonal planes in medical image segmentation is less explored. So, in this study we used non-orthogonal views to see if this can be used as an alternative plane to make the contouring process efficient. But from the result we understood that the use of non-orthogonal planes was cognitively demanding as the users often had to mentally reconstruct the 3D anatomical models. Both participants started with the axial slice and then tried to rotate to get different planes in view. The reason for starting with axial plane might be that both users are used to view the images only in the axial plane. It has been found in the study conducted by Varga et al [VARG2013] that the main strategy of physicians at first is to select an orthogonal slice which they are familiar with as a basis and then rotate to an oblique slice.

Contouring

When compared between the pen on tablet and pen on screen, both users found it very intuitive to contour using pen on screen. Using Prototype 2, it was assumed by the author that the user could visualize the tumour on the screen in any direction in a 2D slice and could easily find the diameter of the tumour to contour on it. However, user 1 started with the same idea but she expected that the design might be similar to the study by Sowell et al. [SOWE2009] where the user could select a largest diameter plane and freeze it and could navigate back and front on the same plane. This type of design might help the users to interpret the anatomy well and could easily modify the contours.

2. HCI devices

Both of our participants are used to contour with mouse and they had no experience in contouring with a pen or finger. As both of the prototypes used pen contouring, the physicians preferred the pen which came with the Wacom tablet as it had a fine tip and it was easy to use. Physicians suggested that they prefer to have many options like contrast adjustments, zooming, panning, etc., which may be controlled by pen, instead of going to the menu on screen and clicking on the screen every time. Both the users also used their fingers on the touch screen during contouring. User 1 who has many years of experience in contouring with a mouse, preferred the concept of using the finger interaction. She mentioned that using finger is a natural way of contouring. But she also mentioned that more testing should be done with the finger interaction, in order to test the accuracy while contouring and also to identify if it is physically or cognitively demanding.

From the results, it could be seen that pen on screen and finger input devices can be a useful input devices in radiotherapy contouring. The use of non-orthogonal view was cognitively demanding hence this information will not be considered in the future UI.

Reflections on the study

Different behaviours were observed when comparing the two participants, which might be related to their experience. The participant who had more experience, was less explorative with the prototypes and to search functions. She was searching for user interface elements which she uses in her daily practice. On the other hand, the less experienced participant was more explorative and found all the functions by herself.

4.2.5 Conclusions

From our study it was clear that interpretation of anatomy with orthogonal plane was very easy because the participants use this setup in their daily routine. Also, the use of pen on screen and finger as an input device were very much appreciated. Hence, it will be considered and explored more in the future studies.

4.3 Design Focus

Based on the prioritized design requirements and the pilot, the design focus, i.e., the crucial improvements to support the interactive segmentation process are:

User input approaches and HCI tools
The influence of user input approaches and HCI tools on the medical image segmentation process and the results are not clear. In the following Chapter 5, we designed and implemented the direct and indirect segmentation approaches with the same workflow but two different types of input tools. Line based tools (contour) will be used with the direct segmentation approach and strokes (area based tool) will be used in the indirect segmentation approach. The aim of Chapter 5 is to investigate the effects of user input approaches and tools in interactive segmentation. Based on the evaluation results, we aim to validate our findings and find new insights on how to further improve the HCI for OAR semi-automatic segmentation procedure.

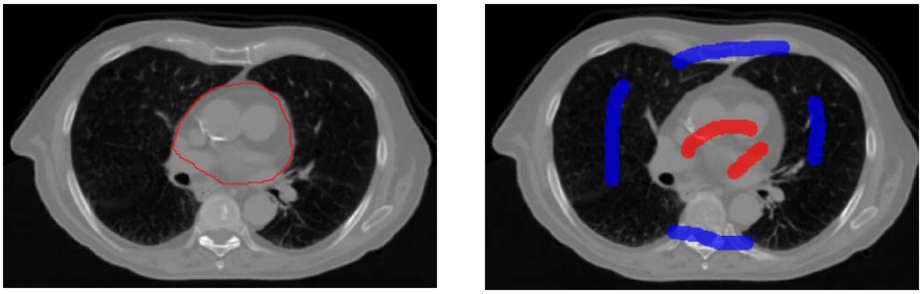
Input devices
The pilot study presented in Section 4.2 reveals the potential of different HCI input devices in interactive segmentation. Together with the outcomes of the literature research and the field study, we will compare the traditional input device design (mouse and pen on pad) and the newly introduced input devices (pen on screen and touch screen) in Chapter 6. Based on the evaluation results of the process and the results, we aim at having a better understanding of the impact of HCI input devices on medical image segmentation and providing choices and design suggestions for further improvements.

Evaluation methods
To evaluate the impact of the design, different types of measures will be used. Chapter 2 summarized many evaluation methods that are used in HCI evaluation for various purpose. Based on Chapter 3, it was found that the heuristic and the think aloud methods could identify general usability issues in the interactive segmentation. However, to have an in-depth understanding the HCI process and its relations to the segmentation results, the performance measure, the physiological measure and the analytical measure, e.g. GOMS, are needed. Besides, the questions of what are the inter-relations between these various HCI evaluation measures in using the input tools and devices, and how to combine the outcomes of those measures to make a proper selection of HCI tools and devices and to identify design issues in their usage remain to be answered. Hence, in Chapter 5 and Chapter 6, various subjective and objective evaluation methods will be incorporated and the

inter-relations among them will be identified. The individual results from the HCI process and the segmentation outcomes and their combined results will be analysed as well to make a proper selection of HCI tools and devices and to identify design issues in the usage.

User input approaches and input tools

5



This chapter is based on the paper:

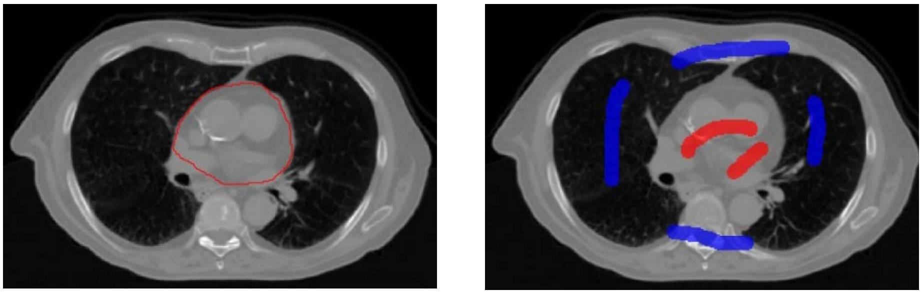
Anjana Ramkumar, Jose Dolz, Hortense A. Kirisli, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle, Laurent Massoptier, Edit Varga, Pieter Jan Stappers, Wiro J. Niessen and Yu Song
"User Interaction in Semi-Automatic Segmentation of Organs at Risk: a Case Study in Radiotherapy". Journal of Digital Imaging, 29(2), 264-277

Anjana Ramkumar, Pieter Jan Stappers, Wiro J.Niessen, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle and Yu Song
"Using GOMS and NASA-TLX to evaluate Human Computer Interaction process in interactive segmentation". International Journal of Human-Computer Interaction, Vol 33(2), 123-134, 2017 .

From Chapter 2 and 3 we were able to identify the gaps in HCI and the improvements that are needed. In Chapter 4, we prioritized our requirements and set the design focus on the user input approaches, tools and input devices. In Chapter 5 and 6, the leading researcher designed the prototypes which utilized different user input approaches, tools and devices. Experienced physicians joined experiments to evaluate the designed prototypes. Based on the evaluation results, we were able to validate our findings and find new insights on how to further improve the HCI for OAR semi-automatic segmentation procedure. The aim of this chapter is to investigate the effects of user input approaches and tools in interactive segmentation in order to propose suggestions for further improvements. This chapter is organized as follows: Two interactive segmentation methods and their prototype design used in this research are introduced in Section 5.1. Section 5.2 lists protocol of the experiment and experimental setup is discussed in Section 5.3. Experimental results are analysed and presented in Section 5.4. The findings in those results are discussed in Section 5.5 where suggestions for the design of user interactions are presented as well. Finally, conclusions are drawn in Section 5.6.

5.1 Prototype Design

In the proposed research, two interactive segmentation approaches with the same workflow but different tools were developed. The first method, which is referred as the “contour” approach, requires the physician to draw contours in a limited number of slices as shown in Fig. 5.1a and the algorithm then computes the segmented volume in 3D. The contour approach is the most familiar method for users, as it is used in many types of clinical software (e.g. Artiview®, 2016; Eclipses®, 2016)[ARTI2016, ECLI2016]. Using this method, physicians were instructed to trace the boundary of the organ accurately on the slice they selected. It is assumed that the interaction can be physically and mentally demanding for the physician. In this context physical demand refers to the laborious and time-consuming contouring. Mental demand refers to the task which involves considerable thinking and scrolling, in which the physician needs to be more focused.

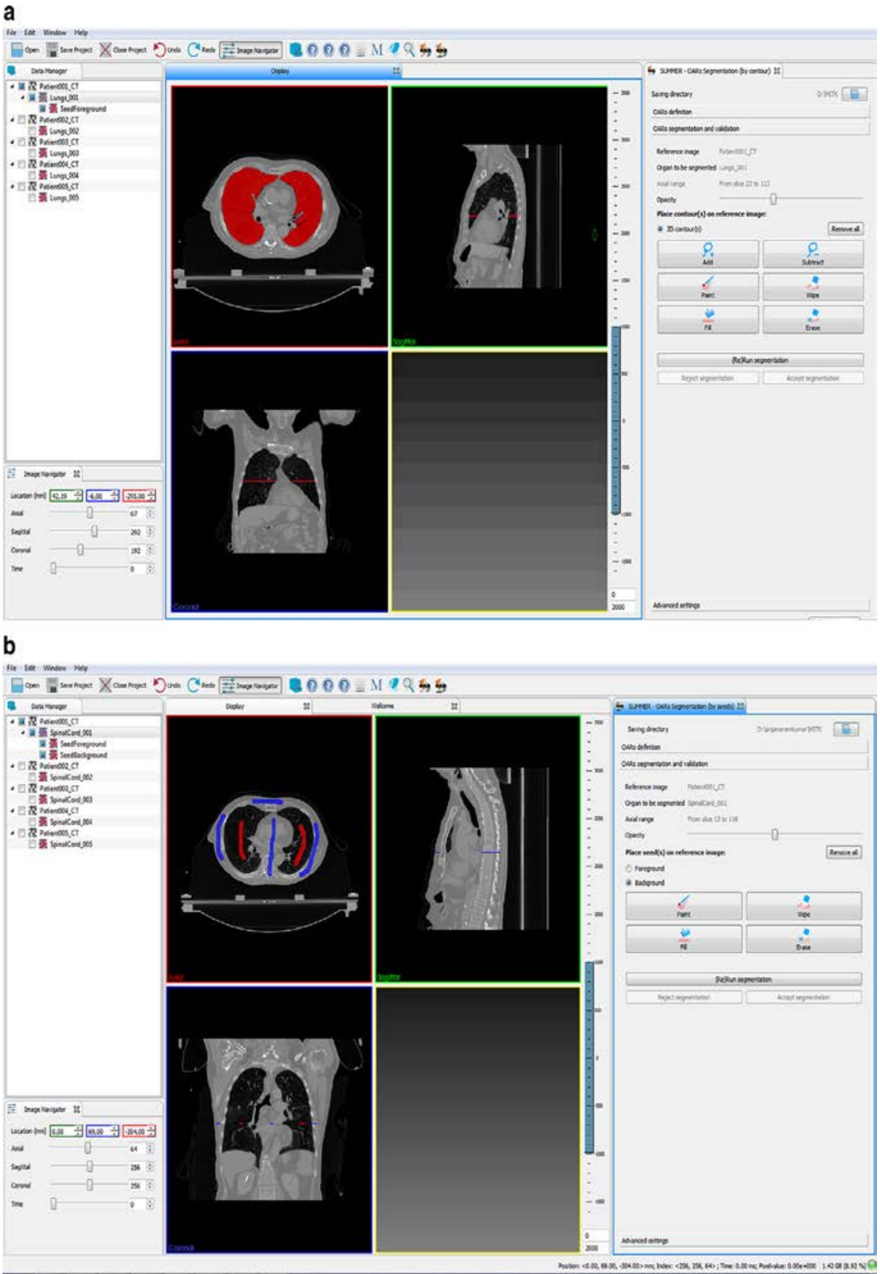


(a) The *contour* approach (b) The *strokes* approach

Figure 5.1: Two designed interactive segmentation approaches

The second interactive segmentation approach is the “strokes” approach which is designed to reduce the physical and mental demands of physicians. The physician draws strokes to indicate the foreground(FG) (as the two red strokes in Fig. 5.1b) that represents the region the physician wants to include as an organ and the background(BG) strokes (as the four blue strokes in Fig. 5.1b) that distinguishes the areas which should not be included in the organ contour. The algorithm then computes the segmentation volume. With strokes interaction, physicians may indicate the region of interest by drawing a line or placing some dots, and it is expected that the physical and mental demands are lower than using the contour method. However, compared to contour method, strokes method is not widely used in radiotherapy.

In order to make a valid comparison of the effects of user interactions in using interactive segmentation methods, the second type of workflow presented in chapter 2(Fig2.2, Intermediate correction) was adopted in both methods as Fig. 5.2a and 5.2b. The reason for using this workflow is to maximally preserve the combined effects of HCI and the algorithm. If manual modifications were allowed, then the quality of the outcome would be hard to judge, as it would be unclear whether it was produced by the interactive segmentation method or manual modifications. In the workflow, after the physician loads a new dataset, he/she can choose either the contour or the strokes method to segment the organ.



5.2a User Interface of the contour method 5.2b User Interface of the strokes method

Figure 5.2: User interfaces of the proposed two interactive segmentation methods.

Physicians can perform actions on axial, sagittal, or coronal planes with the help of HCI tools. The physician may scroll through all the slices, provide certain input on the desired slices and modify until a satisfied input for the algorithm is achieved.

Then the physician runs the algorithm with the provided input and evaluates the outcome. If the outcome is not satisfactory, the physician may re-define the inputs of the algorithm and re-run the segmentation process. Maximally five iterations for each organ were given to the physician and if the result is not satisfactory after the fifth iteration, the segmentation is considered to be unsuccessful. Figure 5.3 shows the workflow of the proposed interactive segmentation method.

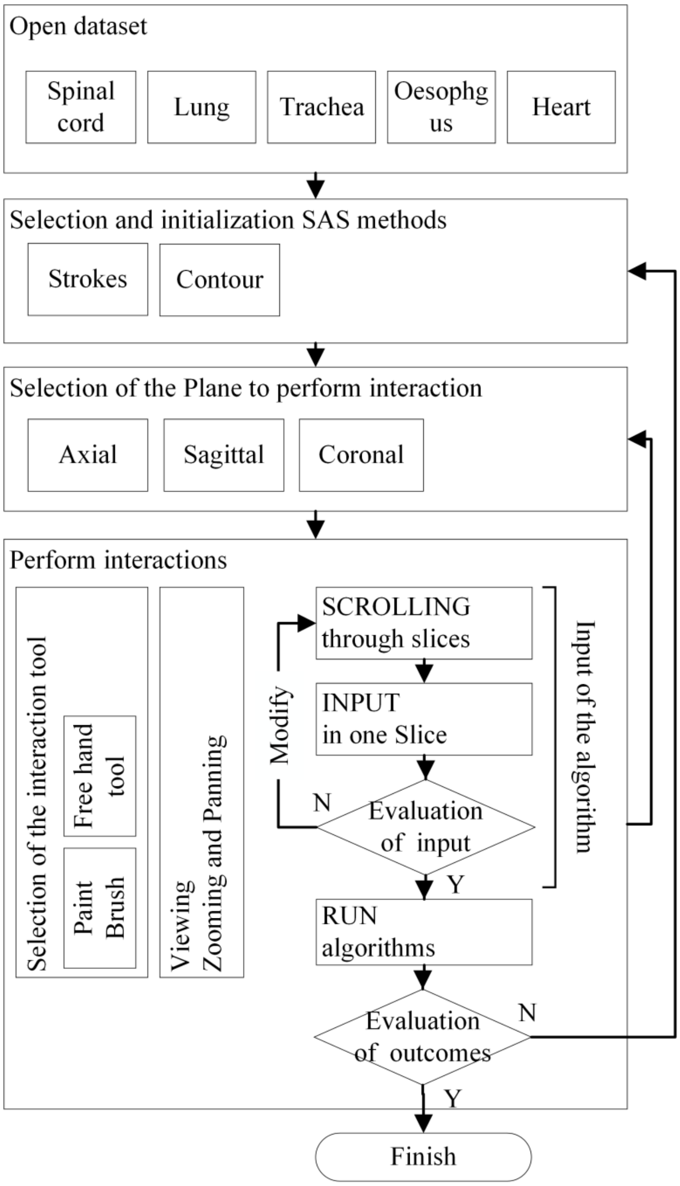


Figure 5.3: Workflow of the proposed interactive segmentation methods

A prototype of both interactive segmentation methods was developed as a plug-in on the medical imaging and interaction toolkit (MITK) platform, version 2013.09.0 [MITK2016]. For both interactive segmentation methods, a combination of graph-cut and watershed-based algorithms was developed by Dolz et al. [DOLZ2014a, DOLZ2014b], and was implemented as the computational part in the prototype. Figure 5.2a and 5.2b shows screenshots of two methods in the prototype. The left window of the display contains the data manager, which allows the physician to select and view the dataset. The main rendering window is presented at the centre with four quadrants, three of them displaying different orthogonal views. The bottom right quadrant shows the segmentation result as a 3D rendering. 2D HCI inputs can be performed in the axial, the coronal, and the sagittal view with a mouse. Tools which can be used for drawing and modifications are on the right side of the interface. In the contour method, a “free hand” tool can be selected by clicking the “add” button on the interface. Besides, physicians can also use a “paint” (paintbrush) tool, with adjustable brush size. In the strokes method, the accuracy requirement of the interaction is not high, thus the “paint” was the only tool that was provided.

Similar to the prototype developed by Heckel et al. [HECK2013], the prototype used in this research is designed in such a way that physicians can give their inputs in any orthogonal planes. Currently in clinical practice, physicians often use only axial view to give their inputs and the other views are often used to check if the segmentation result is satisfactory. By giving the freedom to draw in any orthogonal planes, physicians may choose the plane which requires few HCI. For instance, when segmenting the spinal cord, physicians can segment in the sagittal or coronal planes. It is expected that this design may reduce the number of user inputs, as well as the time taken for drawing the contours/strokes due to fewer slices.

5.2 User Testing Protocol

For a better preparation of user testing, a series of evaluations were performed as shown in Fig. 5.4. The evaluation started with functional testing. Functional testing refers to the test of computational algorithms to evaluate their stability and accuracy. Only after a satisfactory functional testing, usability inspection was performed. Problems identified in the usability inspection were also reported to the developers. Once the issues were fixed, a pilot study [RAMK2014] was conducted to: (a) verify the experimental setup and protocols; (b) overcome the learning curve of physicians, especially for using the strokes method and giving input in different orthogonal planes. After testing the protocols, the case studies were performed and measurements regarding the process and result were collected.

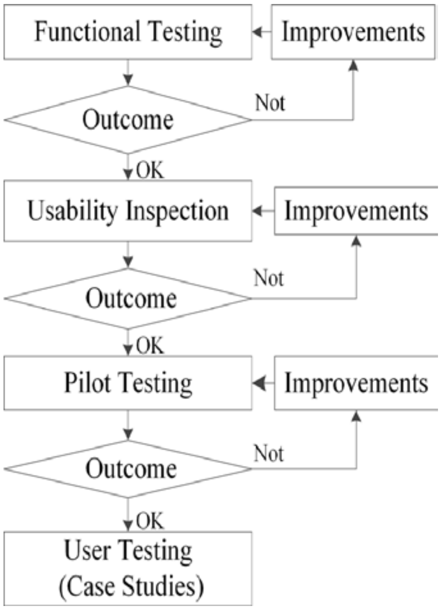


Figure 5.4: The evaluation methods applied in this research

5.3 Experimental Setup

This study was conducted at the Department of Radiation Oncology, The University Medical Centre Freiburg, Freiburg, Germany and Faculty of Industrial Design Engineering, Delft University of Technology, The Netherlands. Datasets of five patients who underwent planning CT (pCT) for lung cancer treatment were selected. Utilization of the datasets for this study was approved by the Ethics committee of The University Medical Centre Freiburg, Freiburg, Germany. Three resident physicians joined the study. The physicians were asked to contour four different types of OAR, i.e., the spinal cord, the lungs, the heart and the trachea using both prototypes, respectively. In the axial direction, the spinal cord and the trachea have a relatively small dimensions where the heart and lungs are larger (diameters of the spinal cord, trachea, heart and right lung in an axial plane are approximately 1-1.5 cm, 2.5 cm, 6.5-7 cm and 12-12.5cm, respectively). Furthermore, the extents of those organs in the sagittal direction (the length) are different. For instance, the spinal cord is approximately 45 cm in length, while the heart is only 12 cm long. Hence the number of 2D CT image slices in the sagittal direction varies as well. Figure 5.5 shows the setup of the study where the prototypes were installed on a laptop. The laptop display (Screen 1) was mirrored on a 22-inch monitor (Screen 2), which is the screen size that physicians are familiar with. A camera was setup in front of the laptop screen to record the complete GOMS analysis and interaction process.GOMS is a specialized human information processor model for HCI observation The software also automatically logged some user interactions into a log file.

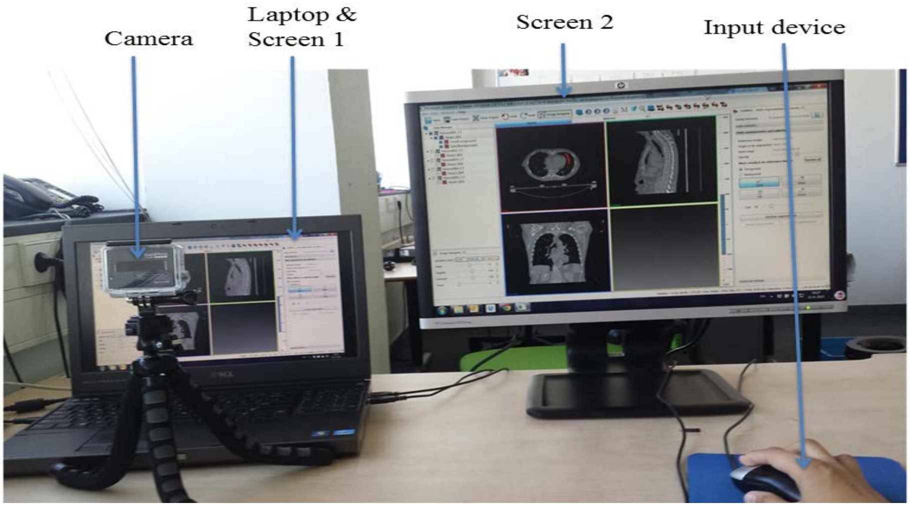


Figure 5.5: The user testing setup

Analytical measure of the process

Based on video analysis, the use of each GOMS [KIER1988] operators in HCI process and its duration are recorded. Apart from this we also measured the number of errors made during the whole segmentation process for each approach. Paired t-tests were used to identify if there are any statistically significant differences among the results.

Subjective measure of the process

In this experiment, the NASA-TLX [HART1988] questionnaire was used as a subjective measure to determine the workload of the user during the segmentation process. Physicians were asked to fill in the questionnaire each time when they finished a case. The questionnaire was divided into six subscale dimensions, each ranging from low to high workload.

5.4 Result

GOMS model

Goals: The top level goal of this task was to segment the organs at risk using two types of user input approaches.

Operators: This study identified mainly 8 categories of operators: *mouse cursor move, zooming, panning, mouse click, scroll, draw* and *brush size adjustment,* that were used in segmenting the OAR. Among them, 5 categories only have one operator and the rest 3, *draw, scroll* and *mouse click*, can be further detailed. The *draw* category has *draw foreground (FG), draw background(BG)* and *draw contour* operators; the *scroll* category also had three operators: *fast scroll, slow scroll* and *normal scroll*; the *click* category consists of five operators: *click FG, click BG, click paint, click add* and *click wipe*. Among those operators, *draw FG, draw BG* and *click FG, click BG* are associated only with the strokes interaction. Table 5.1 shows the operators that were identified in this study for the two types of user inputs. The duration of each operator and the explanation of each operator are presented as well.

Table 5.1:GOMS operators

No	Operators	Time (s)	Meaning
1	*Mouse cursor move*	0.9	Moving of the cursor from the drawing region to a panel to select a tool
2	*Zooming*	2	Right mouse button down and move the mouse
3	*Panning*	2	Mouse middle button down and move the mouse
4	*Mouse clicks* *Click paint* *Click FG tool* *Click BG tool* *Click Add tool* *Click wipe*	0.2	Left mouse button click
5	*Scrolling time* *Slow scroll* *Normal scroll* *Fast scroll*	 0.8 0.3 0.03	Mouse wheel scroll forth and back Observed during decision making process Observed when the user wanted to reach the target region Mainly observed while familiarization with the anatomy of the dataset
6	*Drawing time* *Draw FG*		Left mouse button down Drawing time differed between the organs,

	Draw BG *Draw contour*		interaction methods and physicians. Hence there was no fixed time for drawing
7	*Wipe*	2-6	Left mouse button down Observed mainly when the user created mistakes
8	*Adjustment of the brush size*	0.4-2	Observed with the paint tool, mainly when the users shifted between tools

Methods: In many cases, a fixed combination of multiple operators were used in the HCI process to achieve a certain goal. For instance, click paint was followed by a mouse move and draw operators in order to segment a single slice. Those fixed combination of multiple operators are named methods. Ten different methods were identified in the use of both the input methods as shown in Fig.5.6. In the figure, the vertical axis indicates the operators that were used in the method while the horizontal axis indicates which step this operator was used in the method. At the right of the figure different types of method are explained. In all methods the first two interactions performed by physicians were usually zooming and panning. Hence, zooming and panning operators are not presented in the explanation. The next step which was observed in most of the methods was that physicians chose the tool and started contouring on the presented slice without scrolling to other slices, which indicates physicians' high confidence on the human anatomy. Only in three methods, physicians scrolled to different slices to provide their inputs. The scrolling time and the drawing time of each method may differ due to different numbers of slices scrolled and the dimensions of the organs, respectively.

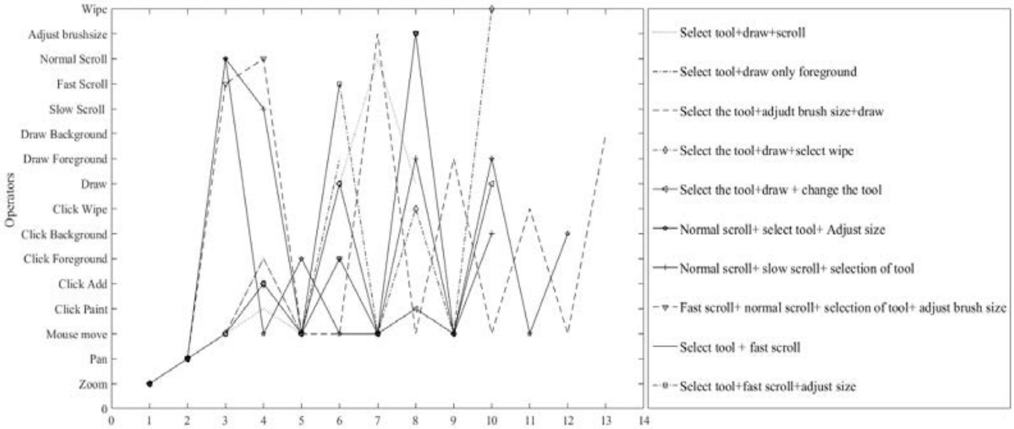


Figure 5.6: Ten different patterns

A workflow is a combination of different methods and operators to achieve a complete segmentation of an organ. The workflow can also be referred as a unit task, as unit tasks refer to the combination of a sequence of smaller tasks in order to achieve a global goal. Figure 5.7 shows two examples of workflows. From Fig.5.7 it can be observed that Workflow 1 is achieved using combinations of method 7 and 1 and Workflow 2 is achieved using method 4 and 1. Using different selection rules, the users may combine different methods together to form different workflows for achieving the same task.

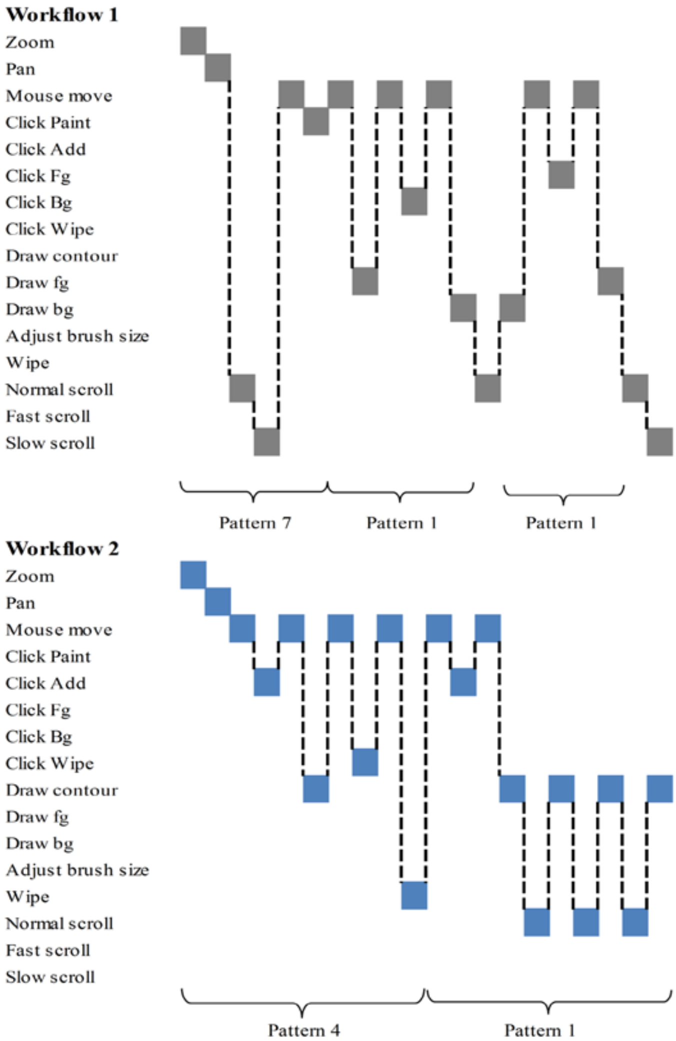


Figure 5.7: Examples of Workflow operators (Workflow 1 is a combination of pattern 7 and 1, Workflow 2 is a combination of pattern 4 and 1)

Total time taken by different Operators

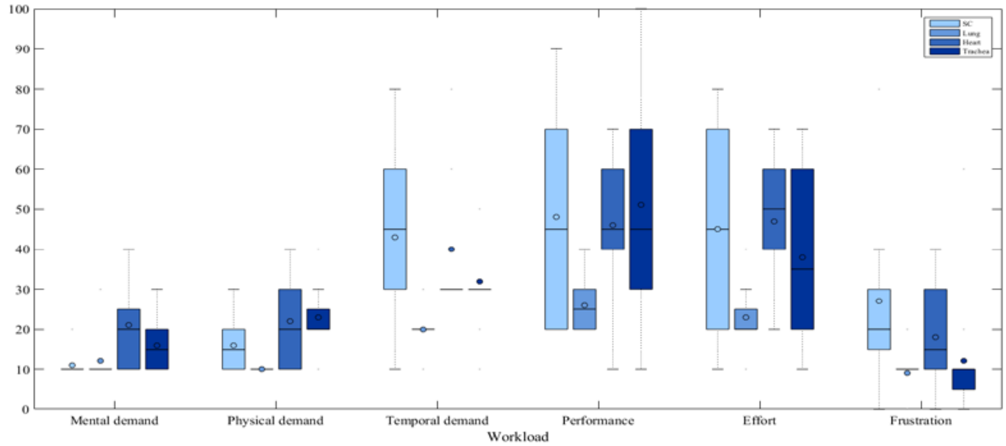
Table 5.2 shows the total time taken by different operators during the whole segmentation process for both interactive segmentation approaches. As mentioned before, the dimensions of the organs are different and hence the overall time of using operators is different for each OAR. However, for each type of operator, except for the drawing time which is strongly associated to the dimensions, the average time taken is nearly same as Table 5.1. When the two approaches were compared against each other, for all the physicians, lung segmentation showed significant difference in the input time (p=0.02) using a paired t-test, where the strokes approach was much faster than the contour approach. Even though there were differences in the mean segmentation time for other organs, these differences were not statistically significant.

Table 5.2: The average time of GOMS operators

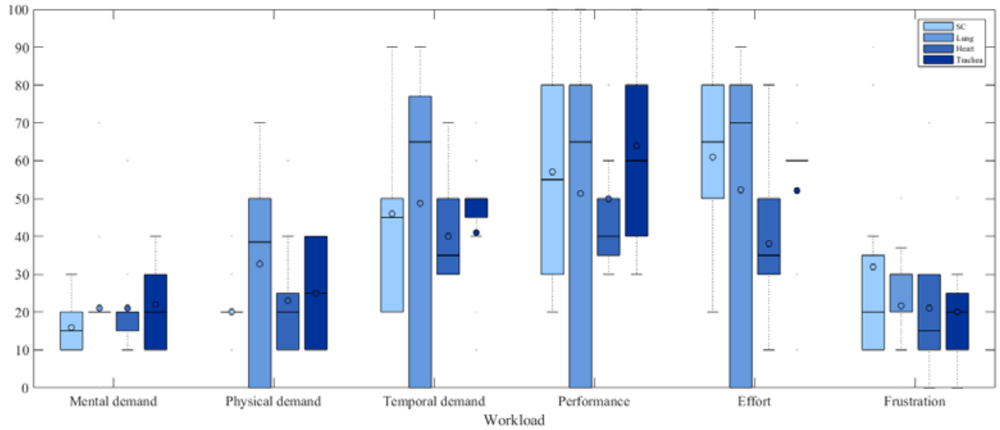
Organs	Type of methods	Drawing time (sec)	Scrolling time (sec)	Normal scroll (sec)	Slow scroll (sec)	Fast scroll (sec)	Mouse moves (sec)	Click time (sec)
Spinal cord	Stroke	35.83±14	36.13±13	20.6±14	7.2±7	4.42	15.9±7	1.3
	Contour	45.28±19	27.33±11.8	18.4±11	2.9±2	3.43±2	13.3±8	0.5
Lungs	Stroke	42.75±7.9	19.24±2	12.5±1	4.2±3	2.52±1	24.3±7	2.4
	Contour	219.75±119	64.04±57	62.5±63	5.8±3	3±2	29.8±20	3±2
Heart	Stroke	65.7±19	19.42±14	15±11	8.8±5	0.4	14.2±8	1.46±1
	Contour	54.78±18	25.64±23	7.4±7	17.4±16	0.7	14.1±10	1.4±1
Trachea	Stroke	51.6±14	13.78±4.5	9.3±3	2.8±3	0	22.14±12	1.68
	Contour	53.8±14	16.86±5.24	6±2	10.8±7	0	12.6±7	1.4

NASA –TLX Questionnaire

Figure 5.8a and 5.8b show the individual workloads for the two types of approaches using NASA-TLX questionnaire. The overall workload is calculated by taking the average of all the individual workloads. The spinal cord and trachea shows higher workload for the contour approach, however the difference was not statistically significant. Only in lung segmentation, a statistically significant difference in the workload (p=0.0002) between the two interactive segmentation approaches was identified.



(a) Using the strokes method



(b) Using the contour method

Figure 5.8: The outcomes of NASA-TLX questionnaires

Dice similarity coefficients of the result

Using the reference standards of each organ, the Dice similarity coefficients of all the organs segmented in Experiment 1 are computed as shown in Table 5.3. P1S indicates Physician 1 using the strokes method and P1C refers to Physician 1 using the contour method. The Dice similarity coefficients of the spinal cord, the lung and the heart are higher than 0.8 in most of the cases.

Table 5.3: Dice similarity coefficient

	Spinal cord				Lung				Heart				Trachea				Oesophagus			
Dataset	**P1S**	**P2S**	**P1C**	**P2C**	**P1S**	**P2S**	**P1C**	**P2C**	**P1S**	**P2S**	**P1C**	**P2C**	**P1S**	**P2S**	**P1C**	**P2C**	**P1S**	**P2S**	**P1C**	**P2C**
Pt 01	0.89	0.87	0.88	0.87	0.97	0.97	0.72	0.97	0.93	0.93	0.93	0.94	0.61	0.62	0.68	0.62	0.75	0.64	0.44	0.29
Pt 02	0.87	0.86	0.87	0.86	0.95	0.95	0.94	0.94	0.90	0.90	0.90	0.91	0.61	0.63	0.68	0.60	0.66	0.68	0.22	0.47
Pt 03	0.84	0.85	0.84	0.26	0.95	0.96	0.96	0.39	0.93	0.93	0.93	0.94	0.57	0.57	0.69	0.33	0.75	0.69	0.49	0.33
Pt 04	0.88	0.88	0.88	0.88	0.98	0.98			0.93	0.93	0.94	0.90	0.71	0.62	0.48	0.54				
Pt 05	0.90	0.88	0.72	0.89	0.98	0.97			0.95	0.92	0.94	0.58	0.63	0.69	0.73	0.66				

Correlations

Table 5.4(a) (b), (c) shows the correlations between measures of the HCI process and performance criteria of the segmentation. For the nine quantitative measures used in the HCI process evaluation, we paired each measure to the others for both types of interaction. A total of 36 pairs were identified for each method. The Pearson correlation coefficient of those pairs are presented in Table 5.4(a) and Table 5.4(b), regarding the contour method and the strokes methods, respectively. Among those pairs of measures, 20 of contour method and 22 of strokes methods were strongly or moderately correlated, either directly or inversely.

Table 5.4(c) shows the correlations of non-quantifiable pairs. A total of ten non-quantifiable pairs were identified for both interactions. The first three pairs are subjective and objective measures in the process and the remaining seven are paired between measures in the process and the result.

(a) The correlations of using the *contour* method. Green: strongly correlated , light green: inversely

For contour	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration	DSC	Drawing time	Scrolling time
Mental demand	1	0.9	0.32	0.08	0.4	0.49	0.01	0.05	-0.12
Physical demand		1	0.42	0.08	0.41	0.44	0.12	0.37	0.08
Temporal demand			1	-0.76	0.89	0.51	-0.39	0.63	0.78
Performance				1	-0.71	-0.32	0.78	-0.35	-0.65
Effort					1	0.73	-0.53	0.4	0.61
Frustration						1	-0.45	0.06	0.28
DSC							1	0.12	-0.23
Drawing time								1	0.84
Scrolling time									1

strongly correlated, orange: moderately correlated and light orange: inversely moderately correlated

For strokes	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration	DSC	Drawing time	Scrolling time
Mental demand	1	0.79	0.32	-0.13	0.44	0.35	-0.25	0.28	-0.04

Physical demand		1	0.59	-0.52	0.59	0.48	-0.45	0.32	0.08
Temporal demand			1	-0.77	0.89	0.58	-0.32	0.67	0.8
Performance				1	-0.86	-0.32	0.47	-0.62	-0.7
Effort					1	0.44	-0.4	0.77	0.8
Frustration						1	0.05	-0.1	0.24
DSC							1	-0.52	-0.18
Drawing time								1	0.78
Scrolling time									1

(b) The correlations of using the *strokes* method . Green: strongly correlated , light green: inversely strongly correlated, orange: moderately correlated and light orange: inversely moderately correlated

Table 5.4: Correlations among different measures in using the *contour* and the *strokes* methods

	SUBJECTIVE AND OBJECTIVE PROCESS AND RESULT correlations		*Strokes*	*Contour*
1	Physical demand	Use of tools	Not applicable because only one tool was used	Not related
2	Effort	Use of tools	Not applicable	Not related
3	Drawing time	Use of tools	Not applicable	Related. Use of paint tool took 3-4 seconds lesser time than free hand tool.
4	Physical demand	Subjective preference	Spinal cord and lung segmentation were directly correlated for both the physicians	
5	Mental demand	Subjective preference	Lung segmentation was very well correlated for both the physicians and spinal-cord was related	

			for one of them	
6	Temporal demand	Subjective preference	Spinal cord and lung segmentation were directly correlated.	
7	Performance	Subjective preference	Spinal cord and lung segmentation were directly correlated.	
8	Effort	Subjective preference	Spinal cord and lung were directly correlated.	
9	Frustration	Subjective preference	For spinal cord and lung it was inversely correlated	
10	Interaction pattern	Dice similarity coefficient	No correlation	Not applicable

(c) List of correlated measures

Predicting NASA-TLX using GOMS operators

Using the linear regression method, we modelled the relations between the workloads identified using NASA-TLX questionnaires and the overall usage time durations of each GOMS operator. In the linear regression, the overall time durations of each of the six GOMS operators, i.e. *Draw, Slow scroll (SS), Normal Scroll (NS), Fast Scroll (FS), Mouse Move (MM)* and *Mouse Click (CLICK)* were used as predictors, and different types of workloads in the NASA-TLX questionnaire were used as criterion variables. Equation 1 and 2 shows the models regarding the strokes approach and the contour approach, respectively. In the regression, the workloads of each physician measured by the NASA-TLX questionnaires were adjusted to a mean of 50 and the standard deviations for different types of workloads and for every physician were normalized as well.

$$\begin{bmatrix} 0.5 & 1 & -0.1 & -0.7 & -0.1 & 2 & 6.9 \\ 0.5 & 0.2 & 0.06 & -1.1 & -0.08 & -5.3 & 18 \\ 0.7 & -0.1 & 1.1 & 0.1 & -0.1 & -7.2 & 18 \\ 0.6 & -0.6 & 0.6 & -0.7 & -0.1 & -8.7 & 50 \\ 0.9 & 0.1 & 0.7 & 1.9 & -0.7 & -8.3 & 20 \\ 0.4 & 0.6 & 0.4 & -0.4 & -0.06 & 4 & 1.3 \end{bmatrix} \begin{bmatrix} DRAW \\ SS \\ NS \\ FS \\ MM \\ CLICK \\ 1 \end{bmatrix} = \begin{bmatrix} MENTALDEMAND \\ PHYSICALDEMAND \\ TEMPORALDEMAND \\ PERFORMANCE \\ EFFORT \\ FRUSTRATION \end{bmatrix}$$

(1)

$$\begin{bmatrix} 0.1 & -0.2 & -0.2 & -1.5 & -0.3 & -1 & 28 \\ 0.2 & -0.2 & -0.3 & 2.8 & -2.6 & 17 & 30 \\ 0.2 & -0.4 & 0.1 & -3.4 & 0.8 & -19.5 & 60 \\ 0.3 & -0.1 & -0.5 & -6.6 & 2.1 & -30.4 & 71.3 \\ 0.1 & -0.2 & 0.1 & -1.9 & 1.1 & -20.8 & 70 \\ 0.02 & 0.1 & 0.5 & -5 & 0.9 & -20.2 & 38 \end{bmatrix} \begin{bmatrix} DRAW \\ SS \\ NS \\ FS \\ MM \\ CLICK \\ 1 \end{bmatrix} = \begin{bmatrix} MENTALDEMAND \\ PHYSICALDEMAND \\ TEMPORALDEMAND \\ PERFORMANCE \\ EFFORT \\ FRUSTRATION \end{bmatrix}$$

(2)

From the model it can be identified that some predictors contribute significantly to one (or several) types of workloads (criterion variables) in the NASA-TLX questionnaire. For instance, the overall time durations of using the *draw* operator and the *slow scroll* (SS) operator are strongly associated to the mental demands when using either the strokes (significance level: 0.001 and 0.01) or the contour approaches (significance level: 0.03 and 0.02). The overall time durations of using the *draw, mouse click and mouse move* operator are strongly associated with the physical demands (significance level: 0.01, 0.007 and 0.04). The time duration of using the *draw* operator and the *normal scroll* operator are also strongly associated with the temporal demand for strokes approach with significance levels of 0.004 and 0.01, respectively. For performance, effort and frustration, we did not find statistically significantly associated predictors.

Errors

Table 5.5 shows the common errors made by the physicians using the proposed two interactive segmentation approaches. A total of 58 errors were identified where 37 of them happened in using the strokes approach and the rest 21 belong to the contour approach. The most common error in using both the approaches is the wrong selection of tools, which contributes to 57% of the total errors. For instance, when the physicians chose the wipe tool they forgot to change it back to the paint tool. Instead they started giving the input using the same tool. The second most common error was in the selection of tools, with physicians sometimes clicking the same option twice resulting in deselection of the tools.

Table 5.5: Percentage of errors in using both methods

Errors	Percentage of error	
	Strokes	Contours
Paint and Wipe operator-With the Wipe tool the users drew on the image and with the Paint tool the users wiped the contour	33%	24%
Click operator-The tool was selected but the user clicked it again and deselected the tool by mistake	7%	10%
Zoom operator - Wrong zoom operations	12%	2%
Click FG and BG operator - Wrong selection of drawing tools (placed BG seeds instead of FG)	7%	-
Click operator - Users forgot to choose the paint tool option instead just selected the FG option	5%	-

5.5 Discussion

In this section, we discuss the outcomes of the subjective and objective measures, GOMS model and the NASA-TLX questionnaire in the evaluation of two interactive segmentation approaches for radiotherapy. First, we discuss the inter-relations between the GOMS model and the NASA-TLX questionnaire. Then the design suggestions regarding the two interactive segmentation approaches are proposed based on a synthesis of the outcomes of the GOMS model and the NASA-TLX questionnaire. Detailed suggestions, which mainly based on the outcomes of the GOMS model regarding each step of the HCI, are proposed as well.

The Use of Correlated Measures

We correlated the subjective and objective measures used in this study and identified the strong, moderate, and weakly correlated pairs. With the paired combinations, it is possible to identify how much effect the designed user interaction has on the HCI process and result. Also the correlated measures provide insights that can be used in improving user interaction design. For example, based on the correlated measures, it was clear that mental demand, physical demand. and temporal demand are correlated to the efforts in both types of interactions and efforts have a direct correlation with frustration. In the use of the contour method, it was observed that frustration and the Dice similarity coefficient are inversely correlated. Hence, efforts and frustration of the users affect the segmentation outcome, as the Dice coefficients represent the quality of the outcome. Thus in future design, the demands of physicians regarding these two aspects should be as low as possible in order to achieve a satisfactory segmentation procedure.

Inter-relations between GOMS model and NASA-TLX

From Table 1 it can be seen that for both interaction approaches, we identified 8 main categories of GOMS operators where drawing, scrolling and mouse clicks also have different variants. Besides, using NASA-TLX questionnaire we identified the workload of the users in using both approaches. In an earlier study conducted by Gao et al. (GAO2015), there was not a single analytical measure that significantly correlated to the workloads in the NASA-TLX questionnaire. However, in this study we were able to identify some individual operators that contribute significantly to the workloads. According to Miyake et al. (MIYA2001), an integrated objective measure is considered more reliable than using an individual measure. We also identified that using combination of measures predicted the workload better than using individual measures. For instance, it was better to predict the physical workload by combining the measures of draw, NS, click and mouse cursor move operators instead of just predicting using draw or NS operator only. The correlation coefficient between the draw operator and the physical demand is only 30%, however, by combining with other operators, the correlation coefficient rises to 60%.

Using regression analysis, we associated the GOMS operators to the mental, physical and temporal demands which were identified by the NASA-TLX questionnaires. Effort and performance demands could not be predicted well using either the individual or combined GOMS operators. A decrease in drawing time will decrease the workload of the users, which was confirmed by the low levels of physical and mental demand found with NASA-TLX using the strokes approach in lung segmentation. In our study, the performance measure on the NASA-TLX questionnaire include aspects of the HCI process while performing the task and are not just limited to the end result. Even after explaining this to users beforehand, the interviews after completion of the tasks indicated that the performance measure was heavily influenced by the end result instead of the HCI process, especially when the quality of the result differs. This partially explains that performance could not be predicted well using either the individual or combined GOMS operators. Hence, we recommend that in a result oriented task, the outcomes of the performance measure should be carefully analysed.

To categorize different operators, we found that the draw operator is associated with both the physical and mental demands, hence it can be categorized as a semi-cognitive and semi-physical operator and the mouse click can be categorized as physical operator. The slow scroll operator contributed significantly to the mental demand in both scenarios. Based on this we concluded that slow scroll is more a cognitive operator than a physical operator. Unlike the mouse click operator, scroll operators identified in this study do not consist of a single task. Instead, it is a fairly complex unit task which may involve different motor, perceptual, and cognitive operators to build up the context. However, we did not have sufficient measures to clearly distinguish if it is a method or an operator. For instance, as we did not measure any eye-movements hence we could not derive which operator contributes to the perception operators in the CPM-GOMS model.

Design Issues

a. The two designed HCI approaches

Based on the results of the GOMS model (Table 5.2), it can be seen that the designed strokes approach was faster in segmenting lungs. The average drawing and scrolling time by the strokes approach in lung segmentation is almost 75% less than the time taken by the contour approach. For the rest of the organs, there was no statistical significant difference in using both input approaches. However, the strokes approach introduces an increased shifting between the FG and BG tools. Consequently, it led to 7% of the total errors.

These findings can be further confirmed by the results of the NASA-TLX questionnaires, especially regarding the associated demands. Except for lung segmentation, there was no statistically significant difference in the workload between the two approaches (Fig. 5.8). It could be explained that the lung is the largest structure (diameters of the spinal cord, trachea, heart and lung in an axial plane were 2 cm, 2.5 cm, 6.5-7 cm and 12-12.5 cm, respectively). Hence, designing tools that are able to automatically identify the type of organ being segmented and adjust their properties accordingly are recommendations for future designs.

b. Other design Issues

The GOMS model has the advantage that it can model the HCI process in a continuous manner where the NASA-TLX questionnaire can only identify the workload of the HCI process at the end of the study [BRUN2006]. Thus, from GOMS we were able to identify more detailed design issues than from the NASA-TLX questionnaire. From Table 5.1, it can be identified that the time taken for operators such as click and release mouse button is in accordance with the literature [KIER1993]. The operator Mouse cursor move took on average 0.2 seconds, which is less than reported in literature [KIER1993]. This may be explained by differences in the mouse travel distance in the graphical user interface.

Table 5.5 shows that switching between the wipe and the drawing tool contributes to 57% of the errors. This was mainly seen in method 4. The wipe tool was used when a mistake was made or physicians were not satisfied with what they drew. One way to solve this issue could be that integrating opposite functionalities in one tool, e.g. using a “Nudge” tool, where the user can enlarge the contour by pushing contour from inside and using the same tool, the use can shrink the contour by pushing it from outside. This will help to reduce the frequency of changing tools. As a result, the distance of mouse movement and the numbers of mouse clicks will drop, which also save time and will reduce the number of errors.

Three different scroll (slow, normal and fast) operators were identified using GOMS model and it was mainly observed in method 6, 7 and 8. In these three methods, the user scrolled through the dataset either to the start or to the end of the dataset. The slow scroll operator was mainly observed when the physicians were making decisions to choose the right slice to provide their inputs by comparing the anatomy and the contour they drew in the previous slice. Hence the time required for this method is longer than others and it involves a lot of decision making processes. This method was observed mainly in segmenting the heart and the trachea. In the case of heart segmentation, at the start of the procedure the physicians do not have the context from the previous or the next slice, so they have to scroll forth and back in order to check the contours and to take the right decisions about the anatomy/structure. In the case of trachea segmentation, the physicians compared the contours to the previously drawn contour in order to include the cartilage. The design suggestions for these methods are that the system can propose a contour on the current slice by considering the previously drawn contours, or two small windows can be designed to show the previous contoured slice and the next slice to be contoured.

Differences in using the strokes and the contour methods

Most of the correlations were nearly the same for using either the contour or the strokes method. However, there are exceptions. One major difference is that drawing time and the subjective performance measure from the questionnaire are strongly correlated for using the strokes method, but not for contour. Besides, we noticed that the drawing time and efforts are strongly correlated in the use of the strokes method. From Table 5.4 it can be seen that the drawing time is less for the strokes interaction in almost all the cases except for segmenting the oesophagus. This concludes that the strokes method was more efficient and effective than contour method. However, it was mentioned by the physicians during the experiment that the cognitive demand of drawing background strokes were higher than drawing foreground strokes. In some case, this higher cognitive demand shifted their preference from using the stokes method to the contour method.

Different from the study conducted by Yurko et al. [YURK2010], our study did not show a strong correlation between mental demand and performance. From Fig.5.8 it is clear that the frustration level of the contour methods is always higher than the strokes method. Besides, the frustration level and Dice similarity coefficient were inversely correlated in using the contour method. With the inverse correlation and from Table 5.3 it can be seen that, outcomes from the contour method are not as good as the strokes method for all the cases and the mental demand, performance and effort were low in using the strokes method. Hence, strokes can be considered as a preferred interaction in future prototypes.

Limitations

One of the limitations of this study is that only three experts participated in the study. For a specialized domain such as radiation oncology, it is difficult to organize a large number of experts as the required expertise is very specific and a considerable amount of time was required for each physician during the pilot, the main experiments and the interviews, etc. Thus the outcomes from this study are more design suggestions for improvement. Besides for some operators, a more in-depth analysis is needed for a more detailed GOMS model with the help of more measures. This will be considered in our future work.

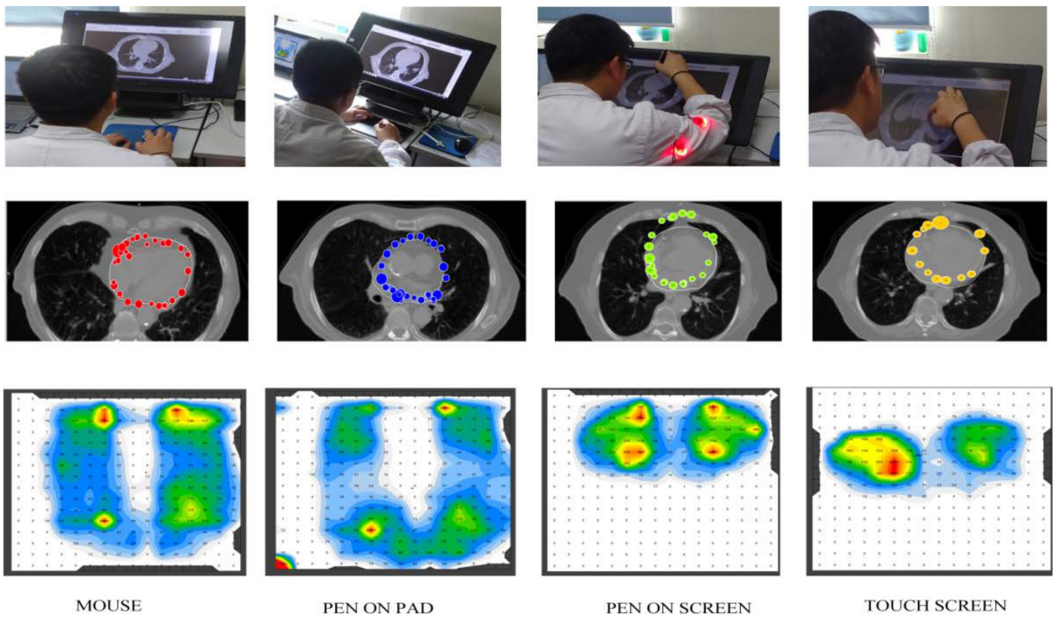
5.6 Conclusions

In this study, we used the GOMS model and the NASA-TLX questionnaire to evaluate the HCI process and to propose design suggestions for interactive segmentation in radiotherapy. Using the GOMS model we identified sixteen different operators and ten different methods that were involved in the segmentation process. Those operators can be further associated to the mental, physical and temporal demands, identified by NASA-TLX questionnaire using regression analysis. The significance of predictors in the regression analysis also helped us identify that if a GOMS operator is a cognitive or physical operator according to its associated demands in the NASA-TLX.

Regarding the segmentation process, the designed strokes approach was faster and less demanding in segmenting large organs based on the findings and inter-relations between the GOMS operators and the results of the NASA-TLX questionnaire. However, it introduces an increased number of shifts between different HCI tools. As a result, physicians tended to make more errors than using the traditional contour approach. For smaller organs, there was no statistical significant difference in using both approaches. Hence, designing tools that automatically identify the organ being segmented and adjust their properties accordingly are recommendations for future designs. Besides, new HCI tools which are able to integrating opposite functions, should be consider as well.

Future study should also focus on involving more HCI components, e.g., new input devices and tools, in order to identify their effects on the HCI process and the segmentation results. More physicians should be involved in the experiment. In addition, more types of subjective, physiological and analytical measures will be incorporated in order to identify the relations among those measurements for offering better design suggestions.

6 HCI Input devices



The design focus of this thesis is to design: 1) user input approaches and HCI tools and 2) HCI input devices to improve the effectiveness and the efficiency in interactive segmentation. In Chapter 5, two user input approaches and different user input tools were designed and evaluated. The aim of this chapter is to have a better understanding of the effects of HCI input devices on medical image segmentation and providing choices and design suggestions for further improvements. In this chapter, Section 6.1 presents the prototype design. The materials and methods used in this chapter are presented in Section 6.2. Section 6.3 shows the test set-up and Section 6.4 presents the evaluation measures used in this study. Section 6.5 summarizes the evaluated results and Section 6.6 discusses the main findings and provides key recommendations for the future. Section 6.7 draws a brief conclusion.

6.1 Prototype Design

Figure 6.1: Designed user interface

To evaluate the effectiveness and efficiency of different HCI devices in medical images segmentation, a software prototype, which is able to utilized the mouse, the pen on pad, the pen on screen and the touch screen as input devices, was developed based on MeVisLab® (version 2.6.2) [MEVI2016]. Figure 6.1 shows a screenshot of the interface of the prototype. On the top part of the interface from right to left are the buttons for saving the contour, deleting the whole contour structure, undo or redoing a drawn contour and selecting the patient case. On the right a slider is provided by which the user is able to navigate through all slices to select a proper 2D slice to work on. In the prototype, only a single contouring tool, by which the users can draw a new contour or modify/delete the existing contour on a 2D slice, was designed to minimize the influences of using different HCI tools.

Regarding different HCI devices, the interface is the same except for the touch screen, the slider was placed on the left. The intention of such design was that physicians can use their left hand to scroll the slices and right hand to draw on the screen. Such a two-hand interaction has both physical and cognitive advantages [LEGA1998]. Figure 6.2 shows the four types input devices used for segmenting the heart (a type of organs-at-risk). Figure 6.2a shows a physician using a mouse as a user input device. The left button on the mouse is used for drawing and the centre button for scrolling. Figure 6.2b shows the physician using a pen on pad. The pad had two buttons on the top left corner using which the physicians could scroll to other slices. Figure 6.2c shows the use of pen on screen. Similar to pen on pad, the scrolling buttons were also incorporated in the interface. Figure 6.2d shows the physician using the touch interface. We designed a slider on the left of the interface for physicians to use it to scroll through different slices.

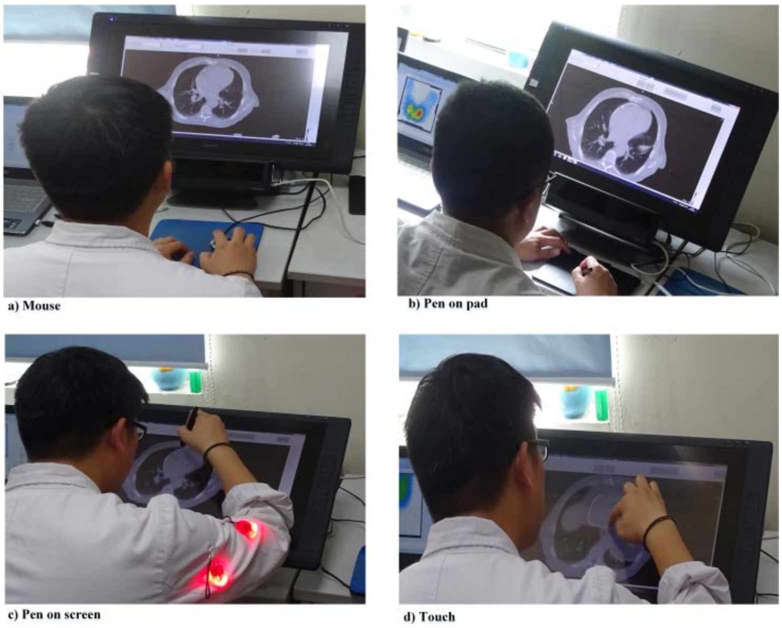


Figure 6.2: Input devices used for segmenting the organ heart

6.2 Materials and Methods

This study was conducted in July and August 2016 at the Department of Radiation Oncology, University Medical Centre Freiburg, Freiburg, Germany, Department of Radiation Oncology, Renji hospital, Shanghai, China and Department of Radiology and Nuclear Medicine, Huashan hospital, Shanghai, China, respectively. A total of 12 Right-handed physicians participated in our study. The experience level of the physicians varied from 2 to 8 years. All physicians had been doing the segmentation task in their clinical routine. Dataset of five patients who underwent planning CT (pCT) for lung cancer treatment were selected. Utilization of the datasets for this study was approved by the Ethics Committee of the University Medical Centre, Freiburg. All physicians participating in the study were informed about the details of the study and signed informed consent forms as well. Before the test, a senior physician was asked to manually segment the organs in each dataset and the outcomes were used as the reference standards.

6.3 Test setup and Protocol

Figure 6.3 shows the experimental setup. In the experiment, the prototype was installed on a laptop (Computer 1). The laptop display (Screen 1) was mirrored on a Wacom® Cintiq® 22HD touch creative pen display (Screen 2) [CINT2016]. A FSA pressure sensing mat [FSA2016] was used to collect physicians seating position. The software for operating the mat was installed in a separate computer (Computer 2). A Tobii X-60 eye-tracker was setup in front of the 22 inch display (~60cm) to capture the eye movements of physicians [TOBI2016]. Two GoPro® cameras were setup: 1) in front of the laptop screen to record the complete interaction process (Camera 1) and 2) beside the physician chair, which recorded the arm movements of physicians during the experiment (Camera 2). Prior to the study, physicians were explained about the designed prototype and input devices. Besides, a 5-10 min session was given to each physician for he/she to get familiarized with the prototype. The task given to the physician was to contour the heart from the CT images using:

1) Pen on screen,

2) Touch / finger input,

3) Pen on pad and

4) Mouse, respectively.

The same sequence was followed for all physicians.

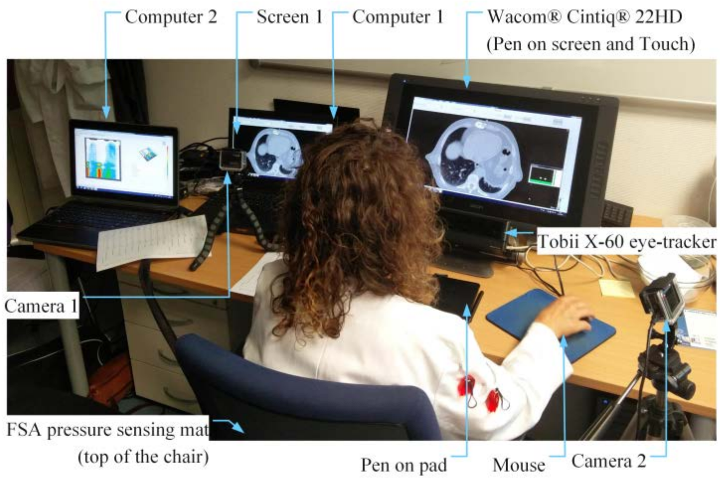


Figure 6.3: Experimental setup

6.4 Evaluation measures

a. Objective measure of the process

Three types of objective measures of the process has been deployed. First, using the video analysis, the efficiency of the segmentation process was measured based on: a) the draw time, b) the scroll time, c) the mouse move time and d) number of corrections required for the users to be satisfied with the outcome. The second part is the eye tracking measure. The average fixation time, the average saccadic movement time and the average pupil diameters were measured using the eye tracker. The final part is the pressure mat measure. The average speed and acceleration of the pressure changes of each physician while doing the segmentation task is calculated based on the recorded pressure maps (0.2 second interval).

b. Subjective measure of the process

In the experiment, each participant was asked to complete a questionnaire. This questionnaire has two parts. First is a device assessment questionnaire suggested by ISO 9241-9 guidelines which consisted of 10 questions [ISO2000]. In order to assess the workload perceived by the physicians in the usage of the device, we also used the NASA-TLX questionnaire as the second part. The NASA-TLX [HART1998] is a self-reported subjective technique for assessing mental workload and was developed by NASA. It consists of a set of six rating scales to evaluate the workload of the users in a task. Those six rating scales are mental demand, physical demand, temporal demand, performance, effort and frustration. Both the questionnaires were combined and made as a single questionnaire and all the questions were scaled from 1 to 10. The questionnaire is presented in Fig.6.4. The right side of the questionnaire indicates the positive value and the left hand relates to the negative side. A bilingual (English-Chinese) expert translated the questionnaire for Chinese physicians.

Figure 6.4: An example of the questionnaire (ISO input device questionnaire and NASA-TLX)

c. Objective measure of the result

The Dice-Jacaard coefficient [DICE1945] was computed to measure the difference between the enclosed areas (2D) /volumes (3D) of the segmentation results and the reference standard. The Dice-Jacaard coefficient can be denoted as $S = 2c/(a + b)$, where a is the volume/area of the segmentation results, b is the volume/area of the reference standard and c is the intersection of a and b.

Directed Hausdorff Distance [HAUS1962] was applied as the basis to measure the deviation between each point on the 2D contours (the outcomes) and the reference stands. DHD delivers the distance from a shape M to another shape E and it can be defined as $H(M, E) = \sup_{r \in M}(\inf_{s \in E}|r - s|)$. In a generalized discreet form, shape M to E will be available as point sets P_M and P_E, where $P_M = \{P_i^M \in M \mid i = 1, m\}$ and $P_E = \{P_i^E \in E \mid i = 1, n\}$, represent M and E, respectively. Thus the DHD from each point in P_M to P_E is $DHD_{ME}^i = \{H(P_i^M, P_E) \mid i = 1, m\}$. In this study, P_E was the reference standard and P_M was one point on the outcome contour. Besides, students t-test was also used to find out if there are any statistically significant differences in the result.

d. Subjective measure of the result

A semi-structured interview was conducted at the end of each session to find out about the personal preference of physicians of the input devices. Besides, physicians were questioned about their opinions of the accuracy of the results and fatigue level in using various input devices.

e. Correlations of measures

To gain additional insights of different measures and propose suggestions based on the synthesis of those measures, we calculated the correlations among different measures of the process and the result using the Pearson product–moment correlation coefficient. These correlations could be (1) correlated; (2) inversely correlated; or (3) not correlated. In this study, we consider p=0.7–0.99 as strongly correlated, 0.4–0.69 as moderately correlated, and 0.1–0.39 as weakly correlated [DANC2004]. Besides, an in-depth analysis of the eye fixations, the physician drawn contour and the gold standard contour regarding each device was also conducted using different measures. An overview of the experiment setup, the participants and different measures is presented in Table 6.1.

Table 6.1: Overview of the experiment setup and the measure

Task	Participant	Input device	Measure	Details of the measures
Segmenting the heart (a type of organs-at-risk)	12 Physicians from 3 hospital	Mouse	Objective measure of the process	Average interaction time Average eye tracking time Average seat pad measures
		Pen on Pad	Subjective measure of the process	NASA-TLX and ISO 9241-9 questionnaire
		Pen on Screen	Objective measure of the result	Dice-Jacaard coefficient and DHD
		Touch screen	Subjective measure of the result	Subjective preference and semi-structured interview

6.5 Results

a. Objective measure of the process

Average Interaction time

Table 6.2: Average interaction time

Activity	Mouse	Pen on pad	Pen on screen	Touch screen
Draw time (sec)	214	186	147	115
Scroll time (sec)	5.42	6.62	6.82	16.82
Mouse move (sec)	0.66	2.56	0.76	0.63
Number of corrections	14	12	11	10

The average drawing time, scrolling time, mouse move time and number of corrections are shown in Table 6.2. When all input devices were compared against each other, there was a significant difference in the drawing time while using the mouse than any other input devices (pen on pad $p=0.02$, pen on screen $p=0.002$, touch screen p=0.002). Also there was a significant difference in drawing time of using touch screen than using pen on pad (p=0.0001) and pen on screen (p=0.01).

Regarding the scrolling time, there was statistical significant difference between using the touch screen and using other input devices (mouse p=0.02, pen on pad p=0.002, pen on screen p=0.004). The mouse move times of the pen on pad and other input devices had a slight difference, but not statistically significant.

Eye tracker

Table 6.3: Average fixation time, average saccadic movement time and average pupil diameter identified by eye tracking

Eye tracker	Mouse	Pen on pad	Pen on screen	Touch screen
Average Fixation time (ms)	286	257	487	446
Average saccadic movement time (ms)	144	123	75	77
Average Pupil diameter (mm)	3.50	3.57	3.55	3.71

Table 6.3 shows the average fixation time, average saccadic movement time and average pupil diameter identified by eye tracking regarding using each input device, respectively. The average fixation time while using the pen on screen and touch screen was higher. Similarly, the average saccadic movement time is lesser with these two input devices.

Seat pressure

Table 6.4: Movements of the users measured by the seat mat

Device	Speed (mmHg/s)	Acceleration (mmHg/s^2)
Mouse	2.42	18.16
Pen on pad	2.14	16.59
Pen on screen	4.15	30.05
Touch screen	5.61	38.41

Table 6.4 shows the movement speed and acceleration of physicians regarding the pressure. When using the touch screen, the average physician's movement speed (5.61mmHg/s) and acceleration (38.41mmHg/s^2) are higher than using other devices, which indicated the usage of the right upper arm. Physicians moved least

while using the pen on pad. We also found that the standard deviation of the average speeds of 12 physicians in using the touch screen (2.9) was higher than other devices (mouse: 2, pen on pad: 1.6, pen on screen 2.7), which indicates a higher inter-observer variability.

b. Subjective measure of the process

Questionnaire

Figure 6.4 shows the ISO 9241-9 and the NASA-TLX questionnaire that was used in this study. The questionnaire had 16 questions, ranging from 0-10. The numbers below 5 represents higher workloads/demands and above 5 was considered as lower workloads or more preferred devices. Figure 6.5 presents the results of the ISO 9241-9 and the NASA-TLX questionnaire. In the figure, it can be seen that the "pen on screen" input device has slightly higher rating than others. When compared between the input devices, it was identified that there are large variations of obscures view, precision/accuracy, general comfort, overall input device, mental effort, physical effort, temporal effort, performance and frustration levels.

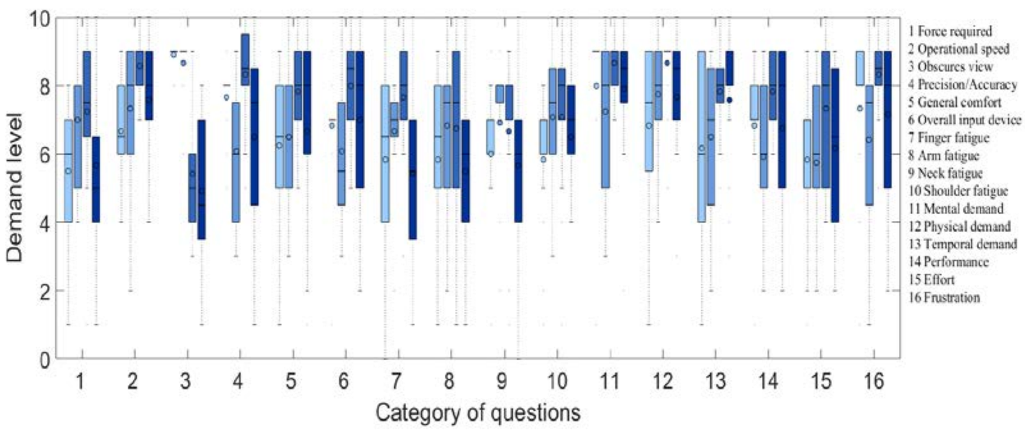


Figure 6.5: The outcomes of ISO- 9241-9 and NASA-TLX questionnaire.

c. Objective measure of the result

The segmentation outcomes of each physician were compared to the reference standards and the Dice Jacaard coefficients were computed. Figure 6.6 summarized the results regarding each input device. The averages of using all four devices are above 0.8. It was also found that the standard deviations of using the mouse, the pen on pad, the pen on screen and the touch screen are 0.03, 0.01, 0.02 and 0.08, respectively. This clearly indicates that:

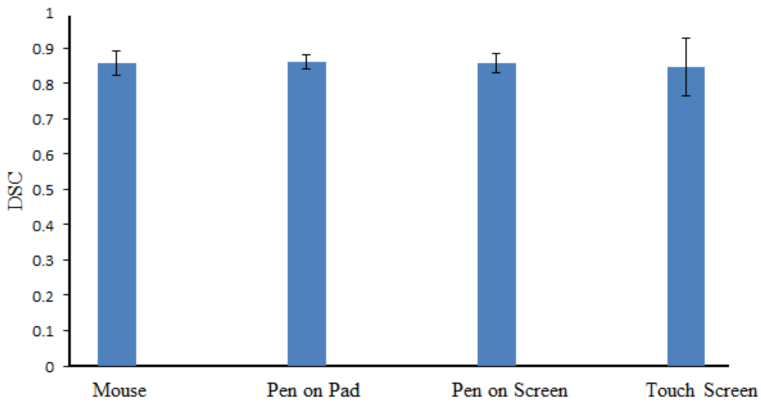


Figure 6.6: Objective measure of the result

1) pen on pad has the least inter observer variability and

2) the touch input device has the larger inter-observer variabilities

d. Subjective measure of the result

Eight out of twelve physicians preferred the "pen on screen" as the input device. Three of them preferred the "pen on pad" and only one suggested the "touch screen". Out of the eight physicians who preferred "pen on screen", four mentioned that "pen on pad" is their second preference. *"Pen on pad needs a bit more practise as it is very sensitive. Maybe after practising I might start liking it as it does not block my view*". Two of the physicians mentioned that, *"the pen on screen blocks a bit of my view but as the organ is very familiar to me, I do not have any problems. Also, I do not have any arm fatigue as I did not do it for a long time"*. Two other physicians who used "pen on screen" mentioned that they would prefer "touch screen" if the scrolling function is move from the interface to physical buttons.

e. correlations

Figure 6.7, 8, 9 and 10 show the correlations between any of the subjective and objective measures of the HCI process and the results of the segmentation regarding four devices, respectively. For the 26 quantitative measures used in the HCI process evaluation, we paired each measure to the others. A total of 325 pairs were identified for each input device. Among those pairs of measures, 150, 120, 120 and 155 measures were strongly or moderately correlated, either directly or inversely for the four different input devices.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1.00	-0.08	0.07	0.10	0.70	0.60	0.83	0.69	0.64	0.78	0.78	0.61	0.45	0.60	0.41	0.47	0.45	0.12	-0.38	0.05	0.37	0.11	0.11	0.29	0.00	0.10	1
	1.00	-0.10	0.13	0.14	0.16	-0.17	-0.19	-0.22	-0.44	-0.05	-0.11	0.33	0.10	-0.06	-0.11	-0.44	0.41	0.45	-0.06	0.12	-0.03	-0.05	0.47	0.68	0.17	2
		1.00	-0.31	-0.34	-0.50	-0.01	-0.29	-0.36	-0.20	-0.34	-0.46	-0.57	-0.46	-0.53	-.627*	-0.06	0.25	0.23	0.61	0.12	0.47	0.51	0.51	-0.02	0.42	3
			1.00	0.42	0.47	-0.14	0.04	0.08	0.01	0.37	0.30	0.23	0.62	0.57	0.69	0.03	-0.46	-0.49	-0.70	-0.28	-0.77	-0.76	-0.02	-0.03	-0.23	4
				1.00	0.82	0.61	0.59	0.51	0.43	0.76	0.82	0.78	0.83	0.68	0.75	0.18	-0.38	-0.61	-0.43	-0.20	-0.26	-0.28	0.04	0.28	0.11	5
					1.00	0.45	0.57	0.54	0.51	0.79	0.79	0.70	0.82	0.76	0.85	0.28	-0.27	-0.55	-0.62	-0.01	-0.53	-0.55	-0.08	0.17	-0.21	6
						1.00	0.89	0.84	0.81	0.70	0.73	0.57	0.47	0.39	0.40	0.28	0.11	-0.35	0.12	0.39	0.29	0.28	-0.04	0.04	0.17	7
							1.00	0.97	0.82	0.72	0.83	0.58	0.51	0.52	0.62	0.14	-0.05	-0.46	-0.17	0.29	0.00	-0.02	-0.30	0.08	-0.05	8
								1.00	0.85	0.73	0.78	0.55	0.50	0.50	0.63	0.16	-0.01	-0.45	-0.16	0.38	-0.01	-0.03	-0.32	-0.02	-0.13	9
									1.00	0.74	0.64	0.38	0.48	0.49	0.55	0.58	0.10	-0.41	-0.03	0.49	0.04	0.04	-0.18	-0.33	-0.18	10
										1.00	0.78	0.61	0.90	0.53	0.76	0.38	-0.14	-0.61	-0.28	0.30	-0.22	-0.23	0.05	-0.15	-0.02	11
											1.00	0.79	0.76	0.78	0.82	0.24	-0.39	-0.63	-0.55	-0.02	-0.34	-0.37	-0.42	0.06	-0.16	12
												1.00	0.72	0.66	0.59	0.02	-0.10	-0.34	-0.41	0.09	-0.14	-0.18	-0.19	0.24	-0.20	13
													1.00	0.65	0.83	0.30	-0.33	-0.65	-0.53	0.05	-0.44	-0.46	0.00	-0.10	-0.13	14
														1.00	0.84	0.41	-0.35	-0.45	-0.78	-0.15	-0.64	-0.66	-0.47	0.04	-0.38	15
															1.00	0.29	-0.51	-0.71	-0.76	-0.17	-0.68	-0.70	-0.35	-0.02	-0.26	16
																1.00	0.07	-0.07	-0.15	0.24	-0.19	-0.18	-0.11	-0.49	-0.07	17
																	1.00	0.78	0.56	0.81	0.51	0.52	0.45	0.12	0.08	18
																		1.00	0.38	0.36	0.33	0.33	0.23	0.27	0.16	19
																			1.00	0.41	0.94	0.95	0.55	-0.07	0.43	20
																				1.00	0.38	0.38	0.24	-0.28	-0.17	21
																					1.00	0.99	0.40	0.03	0.39	22
																						1.00	0.42	0.01	0.40	23
																							1.00	0.24	0.49	24
																								1.00	0.48	25
																									1.00	26

1.Force Required
2.Operation Speed
3.Obscures view
4.PrecisionAccuracy
5.General comfort
6.Overall input device
7.Finger fatigue
8.Arm fatigue
9.Shoulder fatigue
10.Neck fatigue
11.Mental effort
12.Physicial effort
13.Temporal effort
14.Performance
15.Effort
16.Frustration
17.DSC
18.Drawing time
19.Scrolling time
20.Mouse move time
21. Number of corrections
22.Speed
23.Acceleration
24.Fixation time
25.Saccadic movement time
26.Pupil diameter

(150 correlations)

Figure 6.7a: Correlations for the mouse input device. Green: strongly correlated, light green: inversely strongly correlated, orange: moderately correlated and light orange: inversely moderately correlated

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1.00	0.39	0.00	0.28	0.63	0.45	0.63	0.65	0.51	0.45	0.40	0.42	0.53	0.42	0.61	0.59	0.11	0.40	0.31	0.21	0.48	0.29	0.24	0.29	-0.08	0.13	1
	1.00	0.17	0.53	0.59	0.53	0.76	0.44	0.48	0.36	0.26	0.24	0.65	0.57	0.51	0.63	-0.30	0.23	-0.04	-0.25	0.28	0.24	0.23	-0.23	-0.76	-0.52	2
		1.00	0.30	0.08	0.17	-0.11	0.05	0.07	0.18	-0.30	-0.20	-0.07	0.06	0.41	0.02	-0.22	0.55	0.58	0.01	0.22	0.43	0.40	0.38	-0.14	0.06	3
			1.00	0.82	0.86	0.69	0.43	0.51	0.51	0.54	0.34	0.61	0.83	0.74	0.67	0.02	0.31	0.17	0.14	0.24	0.12	0.12	0.05	-0.02	0.10	4
				1.00	0.91	0.88	0.79	0.77	0.55	0.69	0.72	0.84	0.83	0.77	0.82	-0.08	0.29	0.10	0.25	0.19	0.13	0.10	0.08	-0.10	0.03	5
					1.00	0.77	0.64	0.64	0.55	0.68	0.50	0.65	0.85	0.66	0.73	-0.17	0.36	0.26	0.24	0.19	0.27	0.26	0.19	-0.12	0.11	6
						1.00	0.72	0.73	0.65	0.61	0.63	0.78	0.71	0.59	0.74	0.05	0.25	0.03	0.09	0.32	0.21	0.19	-0.06	-0.28	-0.11	7
							1.00	0.96	0.59	0.41	0.83	0.63	0.51	0.43	0.58	-0.07	0.26	0.31	0.39	0.26	0.19	0.12	-0.08	-0.19	-0.17	8
								1.00	0.65	0.41	0.77	0.59	0.55	0.38	0.56	-0.04	0.20	0.31	0.38	0.27	0.24	0.18	-0.24	-0.24	-0.24	9
									1.00	0.15	0.39	0.14	0.26	0.22	0.17	0.49	0.47	0.55	0.30	0.34	0.50	0.45	0.02	-0.12	0.01	10
										1.00	0.46	0.74	0.84	0.52	0.83	-0.25	-0.21	-0.18	0.02	0.01	0.20	0.22	0.15	0.12	0.30	11
											1.00	0.67	0.42	0.31	0.48	-0.05	-0.10	-0.11	0.19	-0.21	-0.16	-0.23	-0.09	0.04	-0.10	12
												1.00	0.83	0.76	0.93	-0.37	-0.07	-0.30	-0.07	0.04	-0.02	-0.04	-0.05	-0.21	-0.13	13
													1.00	0.75	0.94	-0.34	-0.02	0.00	-0.05	0.14	0.18	0.19	0.05	-0.14	0.05	14
														1.00	0.80	-0.18	0.36	0.09	0.06	0.31	0.15	0.14	0.31	-0.01	0.18	15
															1.00	-0.41	-0.02	-0.08	-0.08	0.21	0.18	0.18	0.09	-0.19	0.02	16
																1.00	0.24	0.17	0.38	0.28	-0.11	-0.10	-0.12	0.45	0.25	17
																	1.00	0.66	0.56	0.67	0.35	0.36	0.38	-0.02	0.22	18
																		1.00	0.38	0.57	0.60	0.57	0.41	0.04	0.28	19
																			1.00	0.46	0.00	0.03	-0.04	0.36	0.25	20
																				1.00	0.37	0.42	0.08	-0.02	0.19	21
																					1.00	0.99	0.37	-0.27	0.18	22
																						1.00	0.36	-0.24	0.22	23
																							1.00	0.42	0.77	24
																								1.00	0.83	25
																									1.00	26

1.Force Required
2.Operation Speed
3.Obscures view
4.PrecisionAccuracy
5.General comfort
6.Overall input device
7.Finger fatigue
8.Arm fatigue
9.Shoulder fatigue
10.Neck fatigue
11.Mental effort
12.Physicial effort
13.Temporal effort
14.Performance
15.Effort
16.Frustration
17.DSC
18.Drawing time
19.Scrolling time
20.Mouse move time
21. Number of corrections
22.Speed
23.Acceleration
24.Fixation time
25.Saccadic movement time
26.Pupil diameter

(120 correlations)

Figure 6.7b: Correlations for the pen on pad input device

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1.00	0.63	0.04	0.44	0.52	0.52	0.55	0.82	0.75	0.86	0.48	0.48	0.33	0.56	0.25	0.40	0.38	0.10	-0.41	-0.16	0.16	-0.25	-0.31	0.19	0.30	0.23
2		1.00	0.22	0.63	0.67	0.58	0.09	0.36	0.18	0.31	0.46	0.51	0.23	0.50	0.29	0.61	0.56	0.06	-0.16	-0.25	-0.01	-0.35	-0.40	-0.06	0.02	-0.03
3			1.00	0.26	0.28	0.29	0.28	0.22	0.17	-0.20	0.23	-0.01	0.45	0.37	0.13	0.36	0.00	0.03	0.09	-0.49	-0.14	0.18	0.14	0.00	0.05	0.02
4				1.00	0.77	0.75	0.44	0.51	0.42	0.40	0.56	0.26	0.40	0.77	0.64	0.84	0.46	0.22	-0.11	0.10	0.26	-0.18	-0.21	-0.09	-0.09	-0.10
5					1.00	0.82	0.56	0.63	0.52	0.45	0.48	0.55	0.58	0.86	0.80	0.88	0.53	0.20	-0.18	-0.22	-0.14	0.01	-0.09	-0.03	0.19	0.05
6						1.00	0.61	0.54	0.48	0.52	0.80	0.13	0.26	0.92	0.63	0.90	0.60	0.35	-0.03	-0.29	0.25	0.24	0.19	0.45	0.50	0.47
7							1.00	0.85	0.80	0.71	0.34	0.05	0.48	0.76	0.65	0.49	0.23	0.37	-0.04	-0.27	0.13	0.38	0.32	0.29	0.37	0.32
8								1.00	0.93	0.79	0.34	0.43	0.62	0.72	0.49	0.49	0.16	0.09	-0.43	-0.18	-0.02	-0.05	-0.12	0.06	0.22	0.12
9									1.00	0.79	0.30	0.30	0.49	0.64	0.35	0.38	-0.03	0.08	-0.44	-0.08	0.04	0.02	-0.06	0.18	0.31	0.23
10										1.00	0.40	0.25	0.20	0.58	0.42	0.32	0.41	0.27	-0.27	0.00	0.33	-0.03	-0.07	0.31	0.38	0.34
11											1.00	0.10	0.08	0.62	0.26	0.76	0.52	0.42	0.01	-0.14	0.52	0.10	0.07	0.57	0.56	0.58
12												1.00	0.65	0.22	0.35	0.34	0.29	-0.17	-0.49	-0.02	-0.42	-0.55	-0.65	-0.49	-0.15	-0.38
13													1.00	0.55	0.56	0.41	0.18	-0.27	-0.51	-0.35	-0.51	-0.34	-0.43	-0.51	-0.29	-0.44
14														1.00	0.72	0.82	0.52	0.16	-0.23	-0.39	0.03	0.11	0.05	0.22	0.28	0.25
15															1.00	0.70	0.61	0.36	0.08	-0.05	-0.01	0.17	0.10	-0.14	0.05	-0.08
16																1.00	0.53	0.40	0.02	-0.10	0.19	0.13	0.06	0.18	0.31	0.23
17																	1.00	0.27	0.07	-0.21	0.24	-0.09	-0.10	0.11	0.15	0.12
18																		1.00	0.79	0.21	0.67	0.59	0.56	0.50	0.53	0.52
19																			1.00	0.14	0.48	0.64	0.66	0.32	0.24	0.30
20																				1.00	0.48	-0.09	-0.08	-0.18	-0.18	-0.18
21																					1.00	0.27	0.31	0.55	0.37	0.50
22																						1.00	0.99	0.66	0.69	0.69
23																							1.00	0.68	0.64	0.68
24																								1.00	0.90	0.99
25																									1.00	0.96
26																										1.00

1.Force Required
2.Operation Speed
3.Obscures view
4.PrecisionAccuracy
5.General comfort
6.Overall input device
7.Finger fatigue
8.Arm fatigue
9.Shoulder fatigue
10.Neck fatigue
11.Mental effort
12.Physicial effort
13.Temporal effort
14.Performance
15.Effort
16.Frustration
17.DSC
18.Drawing time
19.Scrolling time
20.Mouse move time
21. Number of corrections
22.Speed
23.Acceleration
24.Fixation time
25.Saccadic movement time
26.Pupil diameter

(120 correlations)

Figure 6.7c: Correlations for the pen on screen input device

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	1.00	0.57	0.80	0.79	0.68	0.63	0.78	0.81	0.74	0.87	0.37	0.65	0.69	0.55	0.60	0.59	0.16	0.30	0.59	-0.22	0.15	0.25	-0.54	0.14	0.38	0.23
2		1.00	0.45	0.68	0.73	0.83	0.66	0.63	0.28	0.31	0.41	0.64	0.81	0.77	0.80	0.83	-0.01	0.05	0.14	-0.52	-0.36	-0.04	-0.11	-0.25	0.07	-0.15
3			1.00	0.67	0.80	0.72	0.66	0.60	0.52	0.67	0.38	0.68	0.72	0.67	0.50	0.66	0.04	0.16	0.44	-0.19	0.14	0.08	-0.55	0.03	0.27	0.11
4				1.00	0.79	0.77	0.83	0.83	0.68	0.70	0.35	0.74	0.86	0.79	0.77	0.76	-0.18	-0.19	0.11	-0.44	-0.11	-0.22	-0.65	-0.24	-0.04	-0.18
5					1.00	0.98	0.75	0.74	0.48	0.59	0.64	0.87	0.81	0.95	0.80	0.95	-0.03	-0.16	0.00	-0.30	-0.06	-0.21	-0.59	-0.11	0.12	-0.04
6						1.00	0.73	0.70	0.40	0.49	0.65	0.83	0.83	0.97	0.87	0.98	0.01	-0.15	-0.03	-0.37	-0.13	-0.24	-0.46	-0.18	0.08	-0.09
7							1.00	0.87	0.73	0.71	0.61	0.63	0.73	0.70	0.71	0.77	-0.18	0.04	0.18	-0.25	0.10	0.08	-0.53	0.19	0.43	0.28
8								1.00	0.83	0.79	0.53	0.70	0.64	0.63	0.76	0.73	-0.05	-0.07	0.13	-0.38	0.02	0.06	-0.50	0.06	0.33	0.15
9									1.00	0.90	0.26	0.58	0.48	0.38	0.51	0.37	0.14	-0.12	0.18	-0.30	0.05	0.07	-0.62	0.13	0.36	0.21
10										1.00	0.38	0.65	0.50	0.48	0.52	0.43	0.24	0.06	0.33	-0.11	0.19	0.05	-0.75	0.20	0.35	0.26
11											1.00	0.44	0.22	0.60	0.64	0.71	0.11	0.03	-0.17	0.12	0.36	-0.18	-0.30	0.31	0.35	0.33
12												1.00	0.77	0.82	0.66	0.76	0.00	-0.39	-0.09	0.55	0.40	-0.30	-0.64	-0.34	-0.02	-0.24
13													1.00	0.84	0.69	0.79	-0.16	-0.14	0.18	0.54	-0.35	-0.11	-0.49	-0.33	-0.01	-0.23
14														1.00	0.84	0.94	-0.03	-0.26	-0.14	-0.34	-0.16	-0.42	-0.56	-0.27	-0.08	-0.21
15															1.00	0.87	0.24	-0.13	-0.06	0.41	-0.06	-0.30	-0.30	-0.20	0.01	-0.13
16																1.00	-0.09	-0.13	-0.07	-0.34	-0.08	-0.21	-0.40	-0.12	0.12	-0.04
17																	1.00	0.33	0.30	0.09	0.20	0.04	0.15	0.11	0.12	0.11
18																		1.00	0.86	0.46	0.57	0.73	0.33	0.62	0.59	0.63
19																			1.00	0.17	0.35	0.73	0.06	0.40	0.51	0.45
20																				1.00	0.80	0.28	-0.06	0.76	0.37	0.65
21																					1.00	0.32	-0.11	0.78	0.50	0.71
22																						1.00	0.30	0.68	0.79	0.74
23																							1.00	-0.04	-0.03	-0.04
24																								1.00	0.87	0.99
25																									1.00	0.94
26																										1.00

1.Force Required
2.Operation Speed
3.Obscures view
4.PrecisionAccuracy
5.General comfort
6.Overall input device
7.Finger fatigue
8.Arm fatigue
9.Shoulder fatigue
10.Neck fatigue
11.Mental effort
12.Physicial effort
13.Temporal effort
14.Performance
15.Effort
16.Frustration
17.DSC
18.Drawing time
19.Scrolling time
20.Mouse move time
21. Number of corrections
22.Speed
23.Acceleration
24.Fixation time
25.Saccadic movement time
26.Pupil diameter

(155 correlations)

Figure 6.7d: Correlations for the touch input device.

Figure 6.8 shows an example of an in-depth analysis of a contouring task using the mouse input device. In the figure, the eye fixation points (cyan), the contour drawn by the physician (black) and the reference standard (red) are presented. It can be seen that the drawn contour and the reference standard were almost completely overlapped except on the left and the top right side. On the left, a “large” saccade can be identified, and the deviation can be associated to this “lost focus”. At the top right, as the number of fixation points are more, it seems that the anatomical structure is difficult and physicians spend more time in identifying it, and such difficulty led to larger variations.

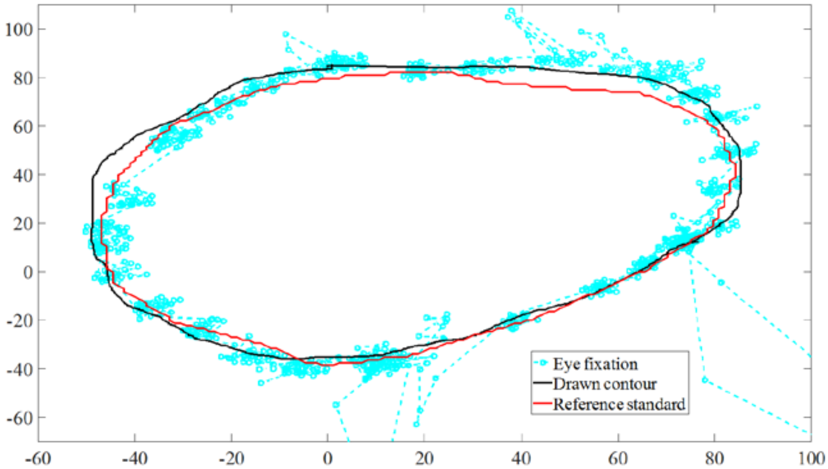


Figure 6.8: An in-depth analysis of the eye fixation versus the objective result for the mouse input device in a single slice.

Figure 6.9 shows the relations between the drawn contour points and the number of eye fixations at a distance of 15 pixels from the contour for all four input devices, respectively. The contour points drawn by the physicians are B-splines which passed a lot of seed points. Generally, the more the "stops" were in the drawing process, the more the seed points were. In the figure, the horizontal axis represents the number of seed points on a physician drawn contour. The vertical axis represents the number of fixation points within 15 pixels of that seed point. It can be found that: 1) using touch screen had the least number of the "stops", which indicated a smooth drawing procedure and 2) the number of fixation points are more while using the mouse.

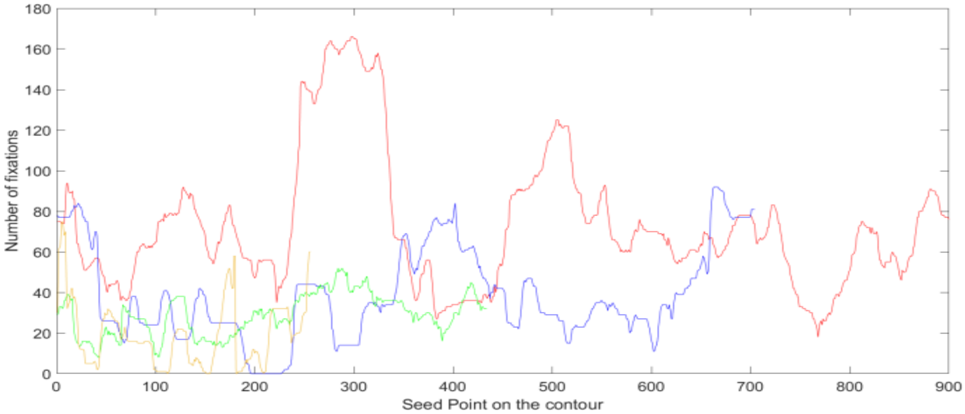


Figure 6.9: Relation between the drawn contour points and the number of eye fixations at a distance of 15 pixels from the contour. Red: Mouse, Blue: Pen on pad, Green: Pen on screen and Brown: Touch screen

6.6 Discussion

In this study, we investigated the impact of using different input devices on the medical image segmentation process and the results using different subjective measures, performance measures, physiological measures and analytical measures. In this section, we discuss the correlations among those measures and based on those correlations, we bring forward suggestions on the selection of HCI devices and interface designs for medical image segmentation.

Correlations

We identified correlations between any two of the 26 measures, either from the process and the results. For using the mouse, we found 150 pairs of correlated measures among 325 possible combinations. For the pen on pad, pen on screen and touch screen, the number of pairs of correlated measures were 120, 120 and 154, respectively. Correlations among measures helped us confirmed the finding and it also provided another way of reasoning those findings.

Using the touch screen as an example, from Fig.6.7d, it can be identified that eye fixations and eye saccadic movements have a strong correlation with the average drawing and the scrolling time. This can be seen in other input devices as well. In touch screen and pen on screen, the seat pad acceleration has a negative correlation with most of the questionnaire categories. Hence, if the speed and the acceleration of the movement of the physicians were large during the task, then most probably they are not satisfied with the input device that they were using. The speed of the movement of the physicians was also inversely proportional to the performance of the input device. The general comfort of physicians has a moderate to strong correlation with the finger, arm, neck and shoulder fatigue. The mouse movement has a negative correlation to the accuracy. More mouse movements in our study often suggested that physicians had to frequently click the delete, undo/redo buttons, which generally indicated that physicians were not satisfied about the outcome and they had change it frequently.

In Fig.6.8, it was found that the number of eye fixation points within a certain distance from the contour were less with the touch and the pen on screen. This can be explained that the physician’s eye moved quickly from one position to another as their view might be blocked by their hands (or pen) while performing the segmentation task. Such findings can be verified from the subjective measure where some physicians mentioned that their view were blocked while segmenting the organs. Besides, from the outcomes of the NASA-TLX it can be found that there are large variations in the obscure view category while using pen on screen and the touch screen. In the figure, it can also be found that the eye fixation points were more using the mouse input. This demonstrated that physicians were not efficient when using the mouse as the movement of the cursor could not be caught by the movement of the eye, which led to many extra fixations. Due to this “slow” movement of the cursor, physicians may lose their concentration, which can be verified from the inter-observer variability of the results as it was larger in using the mouse than using the pen on pad or pen on screen.

Suggestions on the input devices

Though the mouse is the most commonly used input device by physicians in their daily routine, using it cost the longest time to finish a segmentation task. Besides, the number of corrections physicians made was also higher. From the questionnaire, it can be seen that the physical demand was higher as well using the mouse input device, which is confirmed by the high correlation between the drawing time and physical efforts.

On the other hand, the pen on pad, seems to be an efficient tool for segmentation. Based on the shorter average drawing time, the lower inter-observer variability of the results, the lower speed and acceleration of the movement of the body, the lower physical and mental demands, it can be concluded that pen on pad is a preferred input device. However, the average drawing time using the pen on pad and number of corrections required were higher. This is also confirmed from the subjective measure, where several physicians mentioned that they preferred pen on pad but required more practise. As an indirect input device, more practices are often needed to achieve a better hand-eye coordination. Besides, another advantage with pen on pad is that it can be easily used by both left and right handed users.

Many physicians preferred the pen on screen, as they felt that it had a natural way of interaction. Also physicians were efficient in drawing and scrolling using pen on screen. However, from the seat pressure measure (Fig.6.10) it can be found that physicians had to shift their position to the front of the chair, which is not the most optimal sitting position. Besides, during segmentation, the body adjustment speed and acceleration were both high as Table 6.4. Some physicians mentioned that for long term contouring task, for instance contouring a complicated tumour structures, pen on screen might be difficult to use as: 1) it partly blocked the views (Fig.6.10) and 2) there was no proper way to rest the hand except the screen was laid on the table. These findings applied to the touch screen as well. Besides, physicians mainly complained that they wanted to use the physical buttons (buttons on the left of the screen) to achieve the scrolling function instead of the slider on the interface. Though both pen on screen and the touch screen were direct input devices, the inter-observer variabilities of the results were high.

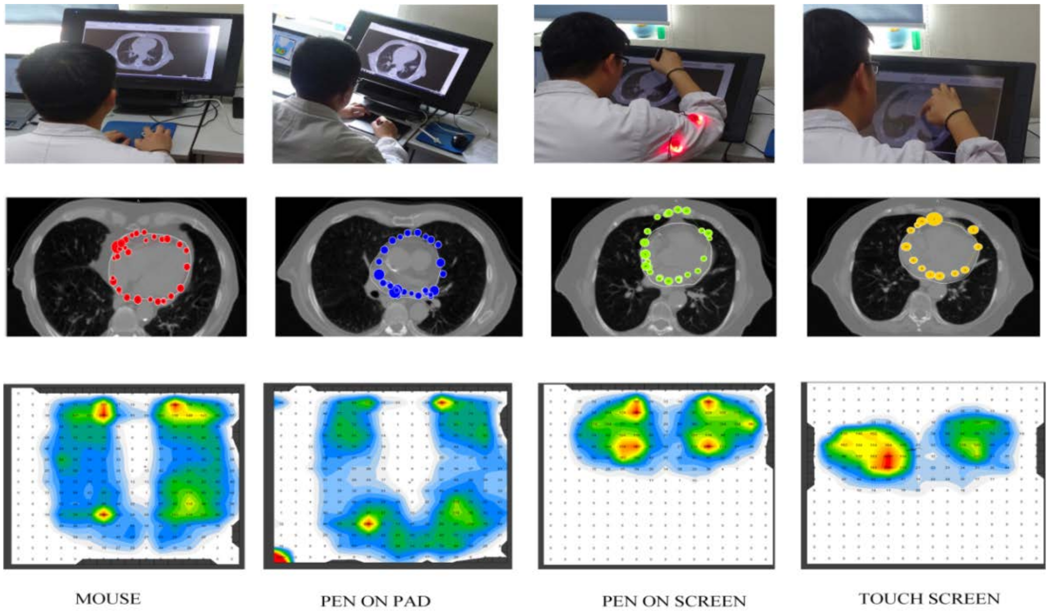


Figure 6.10:Using eye tracking and seat mat to measure the HCI process of using different input devices

Among all the four devices, the muscle involvement is less using the mouse and the pen on pad as it requires only fine motor skills. Also the physicians can sit comfortably while using the mouse and pen on pad input devices. By considering the outcomes of all the measures, it can be seen that pen on pad was the best device for the medical image segmentation, provided that physicians were given more time for practising.

Limitations

One of the limitations of this study is that only 12 radiation oncologists participated in the study. For a specialized domain such as radiation oncology, it is difficult to organize a large number of experts as the required expertise is very specific and a considerable amount time was needed for each oncologist. Besides, current experiments stayed on relatively simple cases, i.e., organs-at-risk. We expected that more can be explored with cognitively challenge tasks, such as contouring tumorous tissues. Finally, due to the limited time available from physicians, we were not able to investigate the fatigue in using different HCI device. This will be considered in our future work.

6.7 Conclusion

In this chapter, we investigated the impact of HCI input devices on medical image segmentation. For this, we used subjective measures, performance measures, physiological measures and analytical measures to evaluate the segmentation process and results with four HCI input devices, i.e., the mouse the pen on pad, the pen on screen and the touch screen. Besides qualitative interviews, we quantified 26 measures regarding each input devices. Among 325 possible combinations, we identified 150, 120, 120 and 154 correlated measures regarding the mouse, the pen on the pad, the pen on screen and the touch screen, respectively. Those correlations indicated the relations among those measures, and also contributed to point out the findings and the reasons behind.

Regarding the segmentation process and result, it was found that using the mouse input was slow and led to many mistakes. Though intuitive, using the pen on screen and the touch screen lead to large inter-observer variabilities among the results since the fingers (pen) blocked the physician's view. Besides, the speed and the acceleration of the movements of physicians were also high, which may cause fatigue in long-term use. The pen on pad was found to be the most suitable input devices for the task. However, as an indirect device, proper training should be provided to physicians to overcome the learning curve.

Discussion & Conclusion

7

This chapter summarizes the outcomes of the research, reflects on the design process and provides an overview of the main contributions of this thesis. Contouring, also referred as delineation or segmentation, is a crucial step in the radiotherapy workflow where organs of interest and the tumour are isolated from the background in order deliver a sufficiently high dose to the tumour while sparing the surrounding normal tissues. The segmentation process can be performed automatically, semi-automatically or manual. Fully automated, operator independent segmentation methods have currently limited applicability owing to the complexity of this task. Patient variability, tumour variability and heterogeneity, and low contrast to noise in medical images makes full automation highly challenging. On the other hand, manual segmentation process is a tedious, time consuming procedure and the quality of the results is prone to inter- and intra- observer variability. Interactive segmentation methods potentially constitute the most promising approaches as a well-designed semi-automatic method is able to combine state of the art image analysis algorithms with physicians' expertise and knowledge to contribute to the effectiveness and efficiency of the segmentation process. To achieve this, efficient methods for computer human interaction (HCI) are needed. However, most literature on interactive automated segmentation do not address this issue.

Hence in this thesis we investigated HCI aspects in radiotherapy contouring, to propose and evaluate effective and efficient HCI design for interactive segmentation. Section 7.1 summarizes the main contributions of this thesis. Section 7.2 reflects on the research methods used and Section 7.3 discusses the limitations of this thesis. Finally, Section 7.4 lists recommendations for the future.

7.1 Contribution of this thesis

The goal of this thesis was to propose effective and efficient HCI designs for the interactive segmentation. First, we performed a literature review on HCI in RT segmentation and we were able to find the answers for the research questions 3-6 that are mentioned in chapter 1. We were able to define what user inputs are most frequently needed during three main phases of the segmentation process: 1) for initializing the computational algorithm; 2) for post-processing corrections and 3) for correcting intermediate segmentation results. We further found that many user input approaches and tools are required to make an efficient and accurate segmentation. We identified four different categories of user input approaches for initializing interactive segmentation, i.e. point based, line based, area based and volume based. Also we determined which input devices are currently used in image segmentation and which of these can potentially be used in assisting interactive segmentation in radiotherapy. As the most recent published review in the HCI was from 2001 [OLAB2001], the review provided in this paper serves as an update on the development of HCI in interactive medical image segmentation. In chapter 3, we identified and presented the workflows (research question 1) and the HCI issues of several popular commercial software solutions that currently are being used for radiotherapy contouring. Using various evaluation methods, we identified several usability and HCI issues and made seven main recommendations (research question 2). These recommendations included that the naming of the buttons should be clear and consistent, automatic segmentation should include necessary and efficient post-processing correction tools, and input device design should synchronize well with the interface design. By using various evaluation measures during the exploration phase, we could bring out that observational and think aloud methods were very useful in identifying the workflow of different systems. The heuristic evaluation method uncovered more peculiar interface design issues.

Based on the conclusions in chapter 2 and 3, we focused the remainder of the research in this thesis on the question how user input approaches and tools and user input devices can be used to use to improve the effectiveness and the efficiency in interactive segmentation. In Chapter 5 and 6, we investigated the impact of user input approaches, tools and input devices on medical image segmentation (research question 7 and 8). In Chapter 5, through an experiment using the designed prototype, we determined the impact of different user input approaches and tools on the interactive segmentation process and result. The task given to physicians during the experiment was to contour four different types of OAR using two different segmentation approaches. We designed and implemented a direct and indirect segmentation approach with the same workflow but with different types of input tools. In the direct approach the physician directly specifies part of the outputs of the delineation task. In the indirect approach, the user roughly specifies the location of the organ, and subsequently the computational algorithm computes the output based on this user provided input. We developed an area based input tool (strokes) and compared it with the traditional line based contouring tool. Among the two developed tools, the area based tool was more efficient, and required less effort than the line based tool. However, it was hard to replace physicians’ subjective preference since cognitively, drawing a contour at the boundary ensures correct segmentation of organs. We also used various subjective, analytical, and performance measures in the evaluation. We identified correlations among those measures and the correlated measures helped us to confirm that besides the performance of the algorithm, the perceived quality of the segmentation also depends on the experience of the physician and the HCI process. Also the findings suggested that in the future HCI design, user interactions need to be less cognitively challenging. The correlations between the subjective and objective measures in this study are also very helpful to identify features relevant for future designs.

In chapter 6 we addressed the question how using different input devices affects the interactive segmentation process and result (research question 8). We designed a prototype which utilized four different HCI input devices, i.e., the mouse, the pen on pad, the pen on screen and the touch screen. We used subjective measures, performance measures, physiological measures and analytical measures to evaluate the segmentation process and segmentation result. Also we identified correlations among and between various measures. Those correlations contributed to identify the design challenges in using various input device in image segmentation.

Regarding the segmentation process and result, it was found that the mouse input was slow and led to many mistakes. The pen on screen and the touch screen input devices were intuitive but led to larger inter-observer variability since the fingers and pen blocked the physician’s view. Besides, the speed and the acceleration of the movements of physicians were high, which may cause fatigue in long-term use. The pen on pad was found to be the most suitable input devices for the task. However, a proper training should be provided to physicians to overcome the learning curve. Hence in future input devices for radiotherapy segmentation an upgrade from the regular mouse input is required.

7.2 Reflection on the research methods

Most likely, the insights obtained in this study could not have been acquired without a co-design approach. To illustrate, at the start of the project, the leading researcher read the HCI literature and did some field research on the existing commercial RT contouring software. However, understanding the HCI during contouring remained challenging and concrete requirements were hard to identify. After applying co-design methods, the contouring process and related needs of physicians became explicit and concrete and with that, we were able to determine the research focus on the user input approaches, tools and input device. By making the radiation oncologist and software developers interact, we were able to have a deeper understanding of the HCI problems and could develop or suggest solutions for effective and efficient HCI for RT contouring.

One of the first methods that we used in order to define our research focus was the observational method. The observational method was very useful during the workflow analysis and identified some minor usability issues. Heuristic, think aloud, semi-structured interviews and questionnaires also helped during the exploration process. The adapted heuristic evaluation method uncovered even more obscure issues with the interface, e.g., design of a button, consistency of a button in two different places etc. With the think aloud method, users were able to compare their system to other contouring systems. Also, it was easier to identify many HCI issues on the contouring tools and on the input devices. The semi-structured interviews with the physicians and the questionnaires allowed to explore deeper level problems and solutions for them. Multiple conversations and observations were required to collect all fragments and to combine them into one complete picture. At the end, a comprehensive overview of the RT contouring workflow and the user involvement in the interactive segmentation was obtained, which helped in prioritizing the design requirements and in-turn helped in the development of the prototype.

Creating prototypes also appeared to be useful during the evaluation phase. By physically testing the prototypes during the evaluation phase, physicians could more easily express their ideas. Providing feedback seemed to become easier and it made physicians more enthusiastic and willing to co-operate, because they were able to really use a working prototype and get a sense of possible future improvements. The prototypes also facilitated the communication between the leading researcher, the physician and the developer by making everyone understand what is needed to make the HCI process efficient and effective.

Physician’s time was the most important issue that had to be taken into consideration during the whole process, especially during the user testing phase. However, all physicians co-operated very well. For the user-testing phase using the prototype, various subjective and objective methods were used. Without using these evaluation methods, it would have been impossible to identify the HCI issues in the segmentation process and the result. Also the correlation between these measures provided useful insights for future user interaction design in interactive segmentation.

7.3 Limitations of this study

Organs: We started our study with the aim of extending our research into HCI issues to tumour contouring. But we could test our design only on relatively simple cases, i.e., OAR. We expect that more HCI issues will be explored with cognitively challenge tasks, such as contouring tumorous tissues.

Different user input tools: From the literature, we identified that there are four different types of user input tools. However, in our study we only compared the line based approach to the area based. The point based and volume based approaches were not explored as it highly depends on the designed computational algorithms. Also, there are many different types of line and area based inputs which might support the segmentation process. Hence, future studies needs to consider involving more tools in interactive segmentation.

Algorithms: Only two combinations of algorithms were used to investigate HCI issues in RT contouring. However, there are other possible algorithms which could be developed in combination with different input tools and approaches. Besides, in Chapter 6, post-processing corrections were done manually due to the lack of programming expertise from the researcher.

7.4 Recommendations for the future

Based on the findings in this thesis, we can provide recommendations both for further research, and for HCI design in novel radiotherapy segmentation software.

With respect to recommendations for further research, this thesis focused only on line and area based initialization of interactive segmentation approaches. In the literature, we identified many alternatives for user input tools. Future studies will need to involve more input tools and approaches that can be used with the input devices, in order to further investigate their effects during the interactive segmentation process.

Additionally, studies will need to focus on designing tools that are cognitively less demanding. As there are research on developing more software tools in HCI, the end user's cognitive demands and evaluation of their cognitive demands while using the tool also needs to be considered. Also, we found that there are no proper tools for objectively measuring the cognitive demands of the users, as most of them are measured as externalised cognition. It is unlikely that only subjective measures are sufficient, and therefore development of objective measures is needed for accurately measuring the cognitive demands of the user in interactive segmentation.

In our studies we found that eye tracking is one of the important evaluation methods to assist in interface design. Still there are no clear recommendations on what are the exact eye tracking measures that can help in the interface design and HCI evaluations. Hence, more research is needed on how to involve eye tracking in HCI evaluation.

In this thesis, we have only focused on OAR segmentation and did not apply our findings to more challenging tasks such as tumour contouring. As the main aim of RT is to deliver a sufficiently high dose to the tumor, while sparing surrounding tissue, accurate and efficient segmentation of tumours is also very important. Owing to tumour heterogeneity, tumour segmentation is more challenging than OAR segmentation, and therefore more likely it will continue to require user interaction. Hence, one of the important directions is to involve tumour segmentation in the research of HCI tools and devices for radiotherapy. Here also, it will be essential to investigate the physical and mental demands of the physician.

In conclusion, much research has focused on algorithms and tools to make segmentation as accurate as possible. However, in practice, corrections are often required, and HCI has received little attention so far. The benefits of combining state of the art algorithms with human expertise are therefore not realized. In order to arrive at accurate and less demanding segmentations in RT, efficient HCI tools are required, for (re-)initialization, and to correct the result at the end of the process. Future studies on radiotherapy should integrate HCI considerations from the start to achieve this. It is expected that when HCI aspects are integrated, the whole segmentation task can be more accurate, fast, and less cognitively demanding.

Summary

Radiotherapy (RT) is the treatment that involves the use of high energy radiations to destroy cancer cells in order to shrink tumours [NHS2016]. Its effectiveness is achieved by damaging the tumour cells' DNA so that these are unable to reproduce themselves. Contouring, also referred to as delineation or segmentation, is an important step in RT workflow where objects of interest are isolated from the background in order to plan a treatment with accurate dose to the tumour and aiming to spare the surrounding tissues.

Literature study indicates that the contouring task is the weakest link in the search for accuracy in radiotherapy. Errors introduced in the contouring task, either from the machine or by human, lead to systematic errors which cannot be eliminated in the subsequent steps. Generally, there are three different ways of performing medical image segmentation: automatic, semi-automatic and manual. Among those methods, semi-automatic methods, i.e., interactive segmentation methods, are potentially the most promising approach as a well-designed semi-automatic method is able to combine the state of the art image analysis with physicians' expertise to contribute to the effectiveness and efficiency of the segmentation method.

Effectiveness and efficiency of an interactive segmentation method depends on the proper combination of physicians' expertise and the capability of the image analysis method. Though physicians play a crucial role in the segmentation process, most of the literature restricted its focus on a specific aspect of the procedure regarding technical elements, such as testing the segmentation algorithm and system accuracy. The cognitive aspects of physicians and HCI in the segmentation process also needs to be given importance and more research needs to be carried out. The goal of this thesis is to

Propose effective and efficient HCI designs for the interactive segmentation.

In order to understand the current interactive segmentation systems and main issues of HCI in using those systems, we reviewed different types of interactive segmentation workflows, user input approaches, tools, and the input devices used for interactive image segmentation tasks in Chapter 2. It was identified that user inputs are often needed during three main phases of the segmentation process: 1) for initializing the computational algorithm; 2) for post-processing corrections and 3) for correcting intermediate segmentation results. Depending on the design of an interactive segmentation method, user input approaches, user input tools and input devices play important roles in (part of) these three phases.

Two types of user input approaches are often used: the direct and the indirect approach, depending on whether the user is asked to directly specify the output of the interactive segmentation method. The direct approach is the most popular approach in current segmentation software solutions, for instance, drawing the contour of the ROI using the pictorial input. However, it could be physically demanding in contouring larger organs such as the lung. Using an indirect approach, the user needs to provide rough, but crucial information to the segmentation algorithms, for instance, the user can roughly specify the locations of the organs for initializing the algorithm. Only a few studies have investigated on the user aspects of the indirect approaches, the effectiveness and efficiency are yet to be explored.

Reviews about user input tools reveal four different categories of tools: point based, line based, area based and volume based tools. However, there are no clear recommendations regarding the influence of these tools on the workflows and what are the differences among them regarding the outcomes.

Different types of input devices are available for radiology and radiotherapy. Besides the regular pen and mouse input, tablet, multi-touch and gesture-based interaction can be used in the contouring task. Thus, exploring the effectiveness and efficiency of using those devices is important to improve the radiotherapy planning process.

Based on the field research, we identified the workflows of existing software solutions and the usability and HCI issues of those software solutions regarding the functions, user satisfactions, limitations, frequently encountered human errors and workloads. Three software solutions were evaluated in different hospitals. The observational research methods, the heuristic evaluation method, the think-aloud and NASA task load index (NASA-TLX) questionnaires were used to get an overview of the workflow and the HCI process of current segmentation systems that are being used in various hospitals for contouring. The observational and think aloud methods were useful in identifying the workflow of different systems. The heuristic evaluation method uncovered more specific issues with the interface design. The think aloud method identified more HCI issues compared to general usability issues. Based on the evaluation results, we were able to identify seven main requirements for the HCI design.

By summarizing the findings in the literature study and field research, we prioritized the requirements as investigating the impact of: 1) different input approaches; 2) different types of pictorial inputs tools and 3) different input devices used in the segmentation process and the results. However, two aspects had to be further confirmed before the we set the design focus: 1) What type of information should be displayed on the interface for segmentation? and 2) How many input devices shall we study? A pilot study was conducted to find answers to the above questions. From the pilot study, we concluded that interpretation of anatomy with orthogonal planes was very easy because the participants use this in their daily routine. Also, the use of pen on screen and finger as an input device were very much appreciated. Hence, we chose to focus on these aspects in our studies. Thus we set our focus in designing: 1) an user input approach and HCI tools and 2) HCI input devices to improve the effectiveness and the efficiency in interactive segmentation.

In Chapter 5 we investigated the effects of user input approaches and tools in interactive segmentation in order to propose suggestions for further improvements. For this, we designed and implemented the direct and indirect segmentation approaches with the same workflow but two different types of input tools. Line based tools (contour) was used with the direct segmentation approach and the strokes, which is an area based tool, was used in the indirect segmentation approach. Two radiation oncologists and a medical physicist joined the experiments to evaluate the designed prototypes. Among the two developed tools, it is clear that the area based tool was more efficient and requires less effort than the line based tool. However it is hard to replace physicians' subjective preference since drawing a contour at the boundary ensures better segmentation of organs. We also used various subjective, analytical, and performance measures during the evaluation. We identified correlations among those measures and the correlated measures helped us to confirm that besides the performance of the algorithm, the quality of the segmentation also depends on the experiences of physicians and the HCI process. Furthermore, the correlated and the inversely correlated measures provide useful insights for future user interaction design in interactive segmentation. Using the analytical measure GOMS, we identified sixteen different operators and ten different methods that were involved in the segmentation process. Those operators were further associated to the mental, physical and temporal demands, identified by NASA-TLX questionnaire using the regression analysis. Besides, it is also identified that random and regular drawing pattern did not influence the quality of the result and the duration of the process. These findings suggest that in the future HCI design of interactive segmentation methods, user interactions need to be less cognitively challenging and there is a need for flexibility in the interface design.

In Chapter 6, we compared the traditional input devices, i.e., the mouse and pen on pad with the newly introduced input devices, i.e., the pen on screen and the touch screen. Based on a software prototype which utilizes four different types of HCI input devices, we evaluated the effectiveness and efficiency of segmentation process and results using different subjective and objective measures in order to 1) understand the relations among those measures and 2) propose design suggestions using the synthesis of these measures. A case study was conducted where twelve physicians segmented OAR using the mouse, the pen on pad, the pen on screen and the touch screen, respectively. We used subjective measures, performance measures, psychophysiological measures and analytical measures to evaluate the segmentation process and results. Besides qualitative interviews, we quantified 26 measures regarding each input devices. Among 325 possible combinations, we identified 150, 120, 120 and 154 correlated measures regarding the mouse, the pen on the pad, the pen on screen and the touch screen, respectively. Those correlations indicated the relations among those measures, and also contributed to point out the findings and the reasons behind. Regarding the segmentation process and result, it was found that using the mouse input was slow and led to many mistakes. Though intuitive, using the pen on screen and the touch screen lead to large inter-observer variabilities among the physicians results since the fingers (pen) blocked the physician's view. Besides, the speed and the acceleration of the movements of physicians were also high, which may cause fatigue in long-term use. The pen on pad was found to be the most suitable input devices for the task. However, as an indirect device, proper training should be provided to physicians to overcome the learning curve. Currently, only a freehand drawing tool was provided to physicians.

With the development of technology, computational algorithms may gradually take over the initial segmentation tasks. We expect that interactive segmentation is moving towards a correction task rather than a creation task. However, the necessity of physicians' review and possible corrections in the post-processing is the rule rather than an exception. Future studies should focus on making the post-processing corrections faster instead of focusing on creating the contour. In the design of the approaches, computer scientists, physicians and designers should closely work together for a feasible solution. More HCI tools and devices should further investigate the effects of using different tools and devices regarding the HCI process and the segmentation results. In addition, more challenging tasks such as tumour contouring needs to be considered for a better understanding of different HCI devices in cognitively demanding tasks.

Samenvatting

Radiotherapie (RT) is de behandeling die omvat het gebruik van hoog energetische straling om kankercellen te vernietigen teneinde tumoren [NHS2016] te doen krimpen. De doeltreffendheid wordt verkregen door het beschadigen van het DNA van de tumorcellen zodat deze zich niet kunnen voortplanten. Contouren, ook wel ‘afbakening’ of ‘segmentatie’ genoemd, is een belangrijke stap in RT workflow waarin objecten van belang (Objects of Interest, OoI) worden onderscheiden van de achtergrond om een behandeling te plannen waarin een nauwkeurige dosis aan de tumor worden uitgevoerd terwijl de omringende weefsels worden gespaard.

Literatuurstudie geeft aan dat het contouren de zwakste schakel is in de zoektocht naar nauwkeurigheid in de radiotherapie. Fouten in de contouren taak, afkomstig van de machine of de mens, leiden tot systematische fouten die niet in de volgende stappen kunnen worden geëlimineerd. Over het algemeen zijn er drie verschillende uitvoeringswijzen voor medische beeldsegmentatie: automatisch, semi-automatisch en handmatig. Onder deze werkwijzen zijn de semi-automatische (d.w.z. interactieve) segmentatiemethoden potentieel de meest veelbelovende aanpak. Dit is omdat een goed ontworpen semi-automatische methode een combinatie kan maken van geavanceerde beeldanalyse met deskundigheid van artsen. Dit draagt bij aan zowel de doeltreffendheid als ook aan de efficiëntie van de segmentatiemethode.

Effectiviteit en efficiëntie van een interactieve segmentatie berust op de juiste combinatie van de expertise van artsen en het vermogen van de beeldanalyse methode. Hoewel artsen een cruciale rol in het segmentatieproces spelen, beperkt de meeste literatuur zich tot een specifiek aspect van de procedure betreffende technische elementen, zoals bijvoorbeeld het testen van het segmentatie algoritme en nauwkeurigheidsbepaling. De cognitieve aspecten van artsen en mens-computer interactie (HCI) in het segmentatieproces moeten ook meegenomen worden, en op dit gebied is meer onderzoek nodig. Het doel van dit proefschrift is

voorstellen ontwikkelen voor effectieve en efficiënte HCI in interactieve segmentatie.

Om de huidige interactieve segmentatie systemen en de belangrijkste vraagstukken van HCI in het gebruik van deze systemen te begrijpen, hebben we in hoofdstuk 2 een inventarisatie en beoordeling gemaakt van verschillende soorten interactieve segmentatie workflows, van methoden voor input van de gebruiker, en van tools en de input devices gebruikt voor interactieve beeld segmentatie taken. Een bevinding was dat de gebruikersinvoer vaak nodig is gedurende de drie belangrijkste fasen van de segmentatie-proces: 1) voor het initialiseren van het computationele algoritme; 2) voor post-processing correcties en 3) voor het bijstellen van tussenresultaten van de segmentatie. Afhankelijk van het ontwerp van de interactieve segmentatie werkwijze spelen gebruikersinvoer benaderingen, gebruikersinvoer gereedschappen, en invoerapparaten een belangrijke rol in (onderdelen van) de drie fasen.

Twee soorten gebruikersinvoer benaderingen worden vaak gebruikt: de directe en indirecte benadering, naargelang de gebruiker gevraagd wordt om de uitvoer van het interactieve segmentatiemethode direct aan te geven. De directe benadering is de populairste benadering in de huidige segmentatiesoftware, bijvoorbeeld door het tekenen van de omtrek van het ROI met de picturale invoer. Deze methode kan echter fysiek zware belasting op de gebruiker zijn bij het aangeven van contouren in grotere organen zoals de longen. Via een indirecte benadering moet de gebruiker ruwe, maar cruciale informatie verschaffen aan de segmentatiealgorithmen. Bijvoorbeeld kan de gebruiker de locaties van de organen ruwweg aangeven om het algorithme te initialiseren. Slechts enkele studies hebben de gebruiker aspecten van de indirecte aanpak onderzocht; met name de doeltreffendheid en de doelmatigheid van de aanpak moeten nog worden onderzocht.

Beoordeling van de gebruikers-invoer onthult vier verschillende categorieën van instrumenten: punt gebaseerde, lijn gebaseerde, gebieds gebaseerde en het volume gebaseerde middelen. Er zijn echter geen duidelijke aanbevelingen gedaan met betrekking tot de invloed van deze gereedschappen op de workflows en over de verschillen tussen de gereedschappen wat betreft de resultaten.

Verschillende soorten invoer apparatan zijn beschikbaar voor radiologie en radiotherapie. Naast de gewone pen en muis invoer, wordt bij de contour-taak gebruik gemaakt van tablet, multi-touch en gebaren gebaseerde interactie. Daarom is het verkennen van de effectiviteit en efficiency van het gebruik van deze apparaten relevant om de radiotherapieplanning werkwijze te verbeteren.

Op basis van het veldonderzoek identificeerden we de werkprocessen van de bestaande software-oplossingen en de bruikbaarheid en HCI-aspecten van deze software oplossingen met betrekking tot de functies, gebruikerstevredenheid, en de beperkingen, zoals vaak voorkomende menselijke fouten en werkbelasting. Drie softwareoplossingen werden geëvalueerd in verschillende ziekenhuizen. Diverse onderzoekmethoden (observatie, heuristische evaluatie, hardopdenkprotocollen en de NASA vragenlijsten over de taakbelasting index NASA-TLX) werden gebruikt om een overzicht van de workflow te krijgen en om het HCI proces van de huidige segmentatie systemen in kaart te brengen. De observatie- en hardop- denken werkwijzen waren bruikbaar bij het identificeren van de workflow van de verschillende systemen. De heuristische evaluatiemethode hielp om meer specifieke problemen in het huidige interface-ontwerp bloot te leggen. De hardop-denken methode identificeerde eerder HCI dan meer algemene usability problemen. Op basis van de evaluatieresultaten konden we zeven belangrijke behoeften met betrekking tot het HCI ontwerp formuleren.

Op basis van de bevindingen van de literatuurstudie en veldonderzoek, kenden we prioriteit toe aan deze eisen, zoals onderzoek naar de impact van: 1) verschillende invoer benaderingen; 2) verschillende typen van picturale invoer gereedschappen en 3) verschillende invoer apparaten gebruikt in het segmentatie proces en de resultaten. Echter, twee aspecten moesten verder worden bepaald voordat het ontwerpfocus kon worden bepaald: 1) Wat voor soort informatie moet worden weergegeven op de interface voor segmentatie? en 2) Hoeveel input devices zullen wij bestuderen? Een pilot-studie werd uitgevoerd om antwoorden op bovenstaande vragen te vinden. Uit de pilot-studie hebben we geconcludeerd dat de interpretatie van de anatomie met orthogonale vlakken erg makkelijk was, omdat de deelnemers dit gebruiken in hun dagelijkse routine. Ook werden het gebruik van de pen op het scherm en de vinger als invoerapparaat zeer gewaardeerd. Vandaar dat we ervoor kozen om ons te concentreren op deze aspecten in onze studies. Derhalve richtten we onze focus in het ontwerpen op: 1) een wijze voor de gebruikersinvoer en de bijgehorende HCI gereedschappen en 2) HCI invoerapparaten om de effectiviteit en de efficiëntie in interactieve segmentatie te verbeteren.

In hoofdstuk 5 onderzochten we de effecten van de verschillende wijzen van gebruikersinvoer en de in interactieve segmentatie gebruikte instrumenten om suggesties voor verdere verbeteringen voor te stellen. Hierbij ontwierpen en implementeerden wij zowel de directe als de indirecte segmentatie benadering met dezelfde workflow maar met twee verschillende invoerapparaten. Voor de directe aanpak werden lijngebaseerde tools (contour) gebruikt; voor de indirecte aanpak oppervlaktebenadering tools (streep). Twee stralingsoncologen en een medisch fysicus namen deel aan de experimenten om de prototypes te evalueren. Uit de vergelijking van de twee ontwikkelde instrumenten werd duidelijk dat de oppervlakte-gebaseerde tool zowel efficiënter als minder belastend was als de lijn gebaseerde tool. Het is echter moeilijk om de persoonlijke voorkeur van artsen te vervangen, omdat het tekenen van een grenscontour zorgt voor een betere segmentatie van organen. In de evaluatie is gebruik gemaakt van diverse subjectieve, analytische en prestatie-indicatoren. We identificeerden correlaties tussen deze indicatorendie aangeven dat de kwaliteit van de segmentatie niet alleen afhankelijk is van de uitvoering van het algoritme, maar ook van de ervaring van artsen en het gevolgde HCI proces. Zowel de positieve als negatieve correlaties leverden nuttige inzichten op voor toekomstige ontwerpen van de gebruikersinteractie bij segmentatie. Met behulp van GOMS identificeerden we zestien verschillende operatoren en tien verschillende methoden die een rol spelen bij het segmentatie proces. Die operatoren werden verder gekoppeld aan de mentale, fysieke en temporele eisen, die met behulp van regressie-analyse verkregen waren uit de NASA-TLX vragenlijst. Daarnaast werd ook vastgesteld dat willekeurig en regelmatig patroon tekening niet de kwaliteit van het resultaat noch de duur van het proces beïnvloedde. Deze bevindingen suggereren dat het toekomstige HCI-ontwerp van interactieve segmentatiemethoden minder cognitief uitdagend moet zijn, en dat er een behoefte is aan meer flexibiliteit in het interface ontwerp.

In hoofdstuk 6 vergeleken wij het traditionele invoerapparaat ontwerp (muis of pen op een invoertablet) met de nieuw ingevoerde input devices (pen op het scherm en aanraakscherm). Op basis van een software prototype dat gebruik maakt van vier verschillende typen van HCI input, evalueerden we de effectiviteit en efficiëntie van het segmentatie proces en de kwaliteit van de resultaten. Hierbij gebruikten we verschillende subjectieve en objectieve maten om 1) inzicht in de relaties tussen deze maten te verkrijgen en 2) een synthese tussen de maten te gebruiken in het voorgestelde ontwerp. Een case studie werd uitgevoerd waarbij twaalf artsen OAR segmentaties uitvoerden met de muis, de pen op pad, de pen op het scherm en het aanraakscherm. Om het segmentatie proces en de resultaten te evalueren gebruikten we subjectieve maatstaven, prestatie-indicatoren, psychofysiologische maten en analytische maten. Naast kwalitatieve interviews, kwantificeerden we 26 maatregelen voor elke input device. Onder de 325 mogelijke combinaties, identificeerden we 150, 120, 120 en 154 gecorreleerd maten ten aanzien van, respectievelijk, de muis, de pen op het pad, de pen op het scherm en het aanraakscherm. Die correlaties wezen op relaties tussen deze maatregelen, en droegen ook bij aan de bevindingen en de achterliggende oorzaken. Bij het segmenteren bleek de muisinvoer traag te zijn en tot veel fouten te leiden. Aangeven met de pen op het scherm en het aanraakscherm werd weliswaar als intuitief ervaren, maar leidde tot grote variabiliteit tussen de gebruikers van de resultaten, waarschijnlijk omdat de vingers en pen het zicht van de arts blokkeerden. Ook vertoonden de bewegingen van de artsen hoge snelheden en versnellingen, wat tot vermoeidheid bij langdurig gebruik kan leiden. De pen op pad bleek de meest geschikte input device voor de taak. Echter, omdat dit een indirecte interactie vergt, moet een goede opleiding aan artsen worden verstrekt zodat zij de ermee gepaarde leercurve kunnen overwinnen. Op dit moment wordt slechts freehand-drawing gereedschap verstrekt aan artsen.

Met de ontwikkeling van de technologie kunnen computationele algoritmen geleidelijk de segmentatie taken overnemen. Wij verwachten dat interactieve segmentatie zich ontwikkelt tot een correctie taak in plaats van een creatie taak. Echter, de noodzaak van beoordeling en eventuele correcties door artsen in de post-processing is nog eerder de regel dan een uitzondering. Toekomstig onderzoek moet er op gericht zijn om het invoeren van post-processing correcties efficienter en makkelijker maken, van in plaats van zich te concentreren op het maken van de contour. In het ontwerp van deze benaderingen is het nodig dat computer wetenschappers, artsen en ontwerpers nauw samenwerken, opdat een haalbare oplossing wordt gevonden. Meer HCI gereedschap en apparatuur moet verder worden onderzocht naar de effecten van het gebruik ervan in het HCI-proces en op de resultaten van de segmentatie. Bovendien moeten meer uitdagende taken, zoals het countouren van worden bestudeerd, om een beter begrip op te bouwen van van verschillende HCI instrumenten in cognitief veeleisende taken.

References

A

[ABDU2011] Abdulin, E. (2011). Using the keystroke-level model for designing user interface on middle-sized touch screens. In: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, pp 673–686

[ALLA2004] Allanson, J., Fairclough, S.H. (2004). A research agenda for physiological computing, Interacting with Computers 16 (5), pp 857-878

[ALON2013] Alonso, R. et al., (2013). Evaluation of head-free eye tracking as an input device for air traffic control. Ergonomics, 56(2), pp 246–55

[ANAL2017] Anatomical plane, Retrieved on Feb.1, 2017, from https://en.wikipedia.org/wiki/Anatomical_plane.

[ANDE2008] Anderlind, E. et al., (2008). Will haptic feedback speed up medical imaging? An application to radiation treatment planning. Acta oncologica (Stockholm, Sweden), 47(1), pp 32–7

[ANDR2014] Andrés, J.D. (2014). Towards an automatic user profiling system for online information sites Identifying demographic determining factors. Online Information Review, 39(1), pp 61 – 80

[ARTI2016]Artiview. http://www.aquilab.com/index.php/artiview.html, accessed 1st April 2016

[ASEL2013] Aselmaa, A., Goossens, R.H., Laprie, A., Ken, S., Fechter, T., Ramkumar, A., Freudenthal, A . (2003). Workflow Analysis Report, http://summerproject.eu/work/deliverables/, accessed 3 May 2016

[AURO2016] http://www.ndigital.com/. (2016). retrived on 1st september 2016.

B

[BACH2015] Bachynskyi, M., Palmas, G., Oulasvirta, A., Steimle, J. and Weinkauf, T. (2015). Performance and Ergonomics of Touch Surfaces: A Comparative Study using Biomechanical Simulation. Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems (CHI2015), 1, pp 1817–1826

[BARR1996] Barret, W.A. and Mortensen, E.N. (1996). Fast, accurate, and reproducible live wire boundary extraction. Vis. Biomedical Computing, pp 183–192

[BAUM2011] Baumgart, D.C. (2011). Smartphones in clinical practice, Medical Education and Research, Arch intern med, 2011; 171 (NO. 14)

[BEIC2012] Beichel, R., Bornik, A., Bauer, C. and Sorantin, E.(2012). Liver segmentation in contrast enhanced CT data using graph cuts and interactive 3D segmentation refinement methods. Medical Physics, 39(3), pp 1361–73

[BELL2010] Bellur, S. and Sundar, S.S. (2010). Psychophysiological responses to media interfaces, To be presented at the proceedings of the 28th conference on Human Factors in Computing Systems (ACM SIGCHI), Atlanta, GA.USA, 2010

[BENK2010]Benko, H. and Wigdor, D. (2010). Imprecision, inaccuracy, and frustration: The tale of touch input. In Tabletops-Horizontal Interactive Displays. Springer, pp 249–275

[BIGD2012] Bigdelou, A., Benz, T., Schwarz, L. and Navab, N. (2012). Simultaneous categorical and spatio-temporal 3d gestures using kinect. In 3D User Interfaces (3DUI), 2012 IEEE Symposium on, pp 53-60

[BLAK2004] Blake, A., Rother, C., Brown, M., Perez, P., and Torr, P. (2004). Interactive image segmentation using an adaptive gmmrf model, European Conference on Computer Vision (ECCV), 428–441

[BRÆN2011] Brændengen, M., Hansson, K., Radu, C., Siegbahn, A., Jacobsson, H. and Glimelius, B. (2011). Delineation of Gross Tumor Volume (GTV) for Radiation Treatment Planning of Locally Advanced Rectal Cancer Using Information From MRI or FDG-PET/CT: A Prospective Study, International Journal of Radiation Oncology*Biology*Physics, 81(4), pp e439-e445.

[BRID2011] Bridger, R. S. and K. Brasher. (2011). Cognitive Task Demands, Self-Control Demands and the Mental Well-Being of Office Workers. Ergonomics, 54 (9), pp 830–839

[BROO1996] Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology, 42, 361–377

[BROO2013] Brooke, J. (2013). SUS : A retrospective. Journal of Usability Studies, 8(2), pp 29–40

[BRUN2002] Bruneau, D., Sasse, M. A., & McCarthy, J. D. (2002). The eyes never lie: The use of eye tracking data in HCI research. In Proceedings of the CHI'02 Workshop on Physiological Computing. NY: ACM Press

[BRUN2006] Bruneau, D.P.J. (2006). Subjective mental workload assessment. International encyclopaedia of ergonomics and human factors, second edition, 1, CRC press, pp 946-947

[BRUN2015] Bruno Garza, J.L. & Young, J.G. (2015). A literature review of the effects of computer input device design on biomechanical loading and musculoskeletal outcomes during computer work. , 52, pp 217–230

[BRYM2006] Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? Qualitative Research, 6(1), pp 97–113

[BRYM2007] Bryman, A. (2007). Barriers to Integrating Quantitative and Qualitative Research. Journal of Mixed Methods Research, 1(1), pp 8–22

[BOUC1992] Boucsein, W. (1992). Electrodermal Activity (New York: Plenum Press)

[BOYK2001] Boykov, Y., Jolly, M.P.(2001). Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. Proc. IEEE Int. Conf. Comput. Vis., pp 105–112

C

[CAIN2007] Cain, B. (2007). A Review of the Mental Workload Literature. Report #RTOTR-HFM-121-part-II, Defense Research and Development, Toronto, Canada, accessed 15th March 2016

[CARD1983] Card, S.K., Newell, A. and Moran. T.P. (1983). The psychology of human-computer interaction. (1983).

[CARM1992] Carmel, E., Crawford, S. and Chen, H. (1992). Browsing in hypertext: A cognitive study. IEEE Transactions on Systems, Man, and Cybernetics, 22(5), pp 865–884

[CHAN2009] Chanel, G., Kierkels, J.J.M., Soleymani, M. and Pun, T. (2009). Short-term emotion assessment in a recall paradigm, International Journal Human-Computer Studies, Vol. 67, pp 607–627

[CHAN2012] Chan, A.J., Islam, M.K., Rosewall, T., Jaffray, D.A, Easty, A.C. and Cafazzo, J.A. (2012). Applying usability heuristics to radiotherapy systems. Radiotherapy and oncology : Journal of the European Society for Therapeutic Radiology and Oncology, 102(1), pp 142–7

[CHEN2011] Chen, J.Y., Seagull, F.J., Nagy, P., Lakhani, P., Melhem, E.R., Siegel, E.L. and Safdar, N. M.(2011). Computer input devices: neutral party or source of significant error in manual lesion segmentation? Journal of digital imaging, 24(1), pp 135–41

[CHAN2016] Chang, C. and Gao, Y. (2016). 3D medical image interaction and segmentation using Kinect. In Biomedical Imaging (ISBI), April 2016 IEEE 13th International Symposium, pp 498-501

[CHRI2012] Christou, G., Ritter, F.E. and Jacob, R.J.K. (2012). Codein-A New Notation for GOMS to Handle Evaluations of Reality-Based Interaction Style Interfaces, International Journal of Human-Computer Interaction, 28(3), pp 189–201

[CINT2016] Cintiq 22HD Touch Creative Pen Display. (2016). Retrieved on Sep.20, 2016 from http://www.wacom.com/en-us/ products/ pen-displays/cintiq-22-hd-touch,

[CORR2015] Corradini, S., Niyazi, M., Niemoeller, O.M., Li, M., Roeder, F., Eckel, R., Schubert-Fritschle, G., Scheithauer, H.R., Harbeck, N., Engel, J. and Belka, C. (2015). Adjuvant radiotherapy after breast conserving surgery – A comparative effectiveness research study, Radiotherapy and Oncology, 144(1), pp 28-34.

D

[DAMA2000] Damany, S. and Bellis, J. (2000). It’s Not Carpal Tunnel Syndrome! RSI Theory and Therapy for Computer Professionals, Jack Bellis, Simax, Philadelphia.

[DANC2004] Dancey, C. and Reidy, J. (2004). Statistics without Maths for Psychology: using SPSS for Windows, London: Prentice Hall

[DANE2006] Danesi, F., Gardan, N. and Gardan, Y. (2006). Collaborative Design: from Concept to Application, Proceedings of the Geometric Modeling and Imaging – New Trends

[DEME1996] Demers, L., Weiss-Lambrou, R. and Ska. B. (1996). Development of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST). Assistive Technology, 8(1), pp 3–13

[DEY2010] Dey, A. and Mann, D.D. (2010). Sensitivity and Diagnosticity of NASA-TLX and Simplified SWAT to Assess the Mental Workload Associated. Journal of Ergonomics, 53(7), pp 848-57

[DICE1945] Dice, L.R. (1945). Measures of the amount of ecologic association between species, Ecology, 26(3), pp 297-302

[DIRI2009] Dirican, A.C. and Göktürk, M. (2009). Attentive Computing, Human Computer Interaction, In Tech Education and Publishing, ISBN: 978-953-7619-16-8

[DIRI2011] Dirican, A.C. and Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in Human Computer Interaction. Procedia Computer Science, 3, pp 1361–1367

[DOLZ2014a] Dolz, J., Kirisli, H.A., Viard, R. and Massoptier, L. (2014). Combining watershed and graph cuts methods to segment organs at risk in radiotherapy, Proc. SPIE 9034, Medical Imaging: Image Processing

[DOLZ2014b] Dolz, J., Kirisli, H.A., Viard, R. and Massoptier, L. (2014). Interactive approach to segment organs at risk in radiotherapy treatment planning, Proc. SPIE 9034,Medical Imaging: Image Processing

[DRUC2002] Drucker, S.M., Glatzer, A., Mar, S.D. and Wong, C. (2002). SmartSkip: consumer level browsing and skipping of digital video content." In Proceedings of the SIGCHI conference on Human factors in computing systems, (ACM, 2002), pp 219-226.

[DUCH2002] Duchowski, A.T. (2002). A breadth-first survey of eye-tracking applications. Behavior research methods, instruments, & computers: a journal of the Psychonomic Society, Inc, 34(4), pp 455–470

[DVOR2014] Dvorak, P., Juras, V., Vogl, W. D., and Chytil, J. (2014). Interactive segmentation of hip joint cartilage. In Progress in Electromagnetics Research Symposium, pp 2369–2372

E

[ECLI2016] Eclipse.(2016). https://www.varian.com/oncology/products/software/treatment-planning/eclipse-proton, accessed on 1st April 2016.

[EGGE2014] Egger, J., Lüddemann, T., Schwarzenberg, R., Freisleben, B., & Nimsky, C. (2014). Interactive-cut: Real-time feedback segmentation for translational research. Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society, 38(4), 285–95

[EHMK2007] Ehmke, C. and Wilson, S. (2007). Identifying web usability problems from eye-tracking data. HCI 2007, pp 119–128

F

[FARM2003] Farmer, E., and Brownson, A. (2003). Review of workload measurement, analysis and interpretation methods, European Organisation for the Safety of Air Navigation: 33 .

[FERY2013] Frey, J., Mühl, C., Lotte, F. and Hachet, M. (2013). Review of the use of electroencephalography as an evaluation method for human-computer interaction. PhyCS - International Conference on Physiological Computing Systems, Lisbonne, Portugal.

[FREU2008] Freudenthal, A., Stüdeli, T.P., Sojar, V., van der Linden, E., Pattynama, P.M.T. (2008). Workflow Development to Allow Early User Guidance of Emerging Image Guidance Technologies, International Conference, Healthcare Systems, Ergonomics and Patient Safety, Strasbourg, France

[FREU2010] Freudenthal, A., Stüdeli, T., Lamata, P. and Samset, E. (2010). Collaborative Co-design of Emerging Multi-technologies for Surgery. Journal of Biomedical Informatics, 44(2), pp 198-215

[FREU2011] Freudenthal, A., Stüdeli, T., Lamata, P. and Samset, E. (2011). Collaborative co-design of emerging multi-technologies for surgery. Journal of Biomedical Informatics, 44(2), pp 198–215

[FROM2011] Fromme, E.K., Kenworthy-Heinige, T. and Hribar, M. (2011). Developing an easy-to-use tablet computer application for assessing patient-reported outcomes in patients with cancer. Supportive care in cancer: official journal of the Multinational Association of Supportive Care in Cancer,19(6), pp 815–22

[FSA2016] FSA pressure mapping. (2016). Vista medical, http://www.pressuremapping.com/userfiles/File/Vista_FSA_overview_for_web.pub_Read-Only.pdf, Retrieved 1st september 2016

G

[GALL2011] Gallo, L., Placitelli, A.P. and Ciampi, M. (2011). Controller-free exploration of medical image data: Experiencing the Kinect. In Computer-based medical systems (CBMS), June 2011, 24th international symposium, pp 1-6

[GAO2012] Gao, Y., Kikinis, R., Bouix, S., Shenton, M., and Tannenbaum, A. (2012). A 3D interactive multi-object segmentation tool using local robust statistics driven active contours. Medical Image Analysis, 16(6), pp 1216–27

[GAO2015] Gao, Q., Yang Wang, Y., Song, F., Li, Z. and Dong, X. (2015). Mental workload measurement for emergency operating procedures in digital nuclear power plants, Ergonomics, 56(7), pp 1070-1085

[GERI2016] Gering, D.T. (2016). System and method of voice activated image segmentation. U.S. Patent 20,160,104,293

[GOLD1999] Goldberg, J. and Kotval, X. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Construct. International Journal of Industrial Ergonomics, 24(6), pp 631–645

[GOME2012] Gomez-Iturriagaa, A., Bilbaoa, P., Casqueroa, F., Cacicedoa, J. and Crook, J. (2012). Smartphones and tablets Reshaping radiation oncologists', reports of practical oncology and radiotherapy; 1 7, pp 276–280

[GOOD2014] Goode, W. et al., (2014). A mixed-method approach for in-depth contextual user research. Lecture Notes in Computer Science (including subseries

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8517 LNCS (PART 1), pp 86-95

[GRAU2014] Grau, C., et al.,(2014). Radiotherapy equipment and departments in the European countries: Final results from the ESTRO-HERO survey, Radiother Oncol 112(2), pp155-64

H

[HARD2003] Harders, M., Member, S., Székely, G., Member A. (2003). Enhancing Human– Computer Interaction in Medical Segmentation, Proceedings of the IEEE, 91(9)

[HART1988] Hart, S.G. and Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in: Hancock, P.A. and Meshkati, N. (Eds.) Human Mental Workload, North-Holland.

[HAUS1962] Hausdorff, F. (1962). Set Theory, 2nd ed., Chelsea, New York, 1962

[HAZL2003] Hazlett, R.L. (2003). Measurement of user frustration: A biologic approach. Proceedings of CHI 2003 Conference on Human Factors in Computing Systems v2, ACM Press, pp 734-735

[HECK2009] Heckel, F., Moltz, J. H., Bornemann, L., Dicken, V., Bauknecht, H.-C., Fabel, M., … Peitgen, H.-O. (2009). 3D contour based local manual correction of tumor segmentation in CT scans. Proceedings of SPIE, 7259, 72593L–72593L–9

[HECK2013] Heckel, F., Moltz, J.H., Tietjen, C., Hahn, H.K. (2013). Sketch-Based Editing Tools for Tumour Segmentation in 3D Medical Images. Computer graphics forum, 32(8), pp 144–157

[HORN2006] Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies, 64(2), pp 79–102

[HÖTK2013] Hötker, A.M., Pitton, M.B., Mildenberger, P., and Düber, C. (2013). Speech and motion control for interventional radiology: requirements and feasibility. International journal of computer assisted radiology and surgery. International Journal of Computer Assisted Radiology and Surgery, April 2013,8, pp 997–1002

[HUNG2016] Hung, H.C and Shelley Shwu-Ching Young, S.S.C. (2016). Applying multi-touch technology to facilitate the learning of art appreciation: from the view of motivation and annotation. Journal of Interactive learning environment. Published online on 9th May 2016.

I

[ISAM2008] Isambert, A., Dhermain, F., Bidault, F., el al. (2008). Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context, Radiotherapy and Oncology, 87(1), pp 93–99

[ISO1998] ISO 9241-11. (1998). Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability.

[ISO2000] ISO. (2000). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. In ISO 9241. International Organization for Standardization, 2000

J

[JACO2003] Jacob, R.J.K. and Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research. Ready to Deliver the Promises. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, pp 531–553

[JAIM2007] Jaimes, A and Sebe, N. (2007). Multimodal human–computer interaction: A survey. Computer Vision and Image Understanding, 108(1-2), pp 116–134.

[JAKO2016] Jakobsen, M.R. and Hornbæk, K. (2016). Negotiating for Space? Collaborative Work Using a Wall Display with Mouse and Touch Input. InProceedings of the 2016 CHI Conference on Human Factors in Computing Systems, (CHI2016), pp 2050-2061.

[JASP2009] Jaspers, M.W.M. (2009). A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. International Journal of Medical Informatics, 78(5), pp.340–353.

[JAUM2014] Jaume-i-Capo, a & Samcovic, A. (2014). Vision-based interaction as an input of serious game for motor rehabilitation. Telecommunications Forum Telfor (TELFOR), 2014 22nd, pp 854–857

[JEON2015] Jeon, M. and Croschere, J. (2015). Sorry, I'm late; I'm not in the mood: Negative emotions lengthen driving time. 12th International Conference on Engineering Psychology and Cognitive Ergonomics, Volume 9174, 2015, pp 237-244

[JOHN1996] John, B.E. and Kieras, D.E. (1996). Using GOMS for User Interface Design and Evaluation: Which Technique? ACM Transactions on Computer-Human Interaction, 3(4), pp 287–319.

K

[KANG2004] Kang, Y., Engelke, K. and Kalender, W. A. (2004). Interactive 3D editing tools for image segmentation. Medical Image Analysis, 8(1), pp 35–46

[KAPT2010] Kaptein, M.C., Nass, C and Markopoulos, P. (2010). Powerful and consistent analysis of Likert-type rating scales. In Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI ’10, page 2391, New York, New York, USA, 2010. ACM Press

[KARA2013] Karasev, P., Kolesov, I., Fritscher, K., Vela, P., Mitchell, P., and Tannenbaum, A. (2013). Interactive medical image segmentation using PDE control of active contours. IEEE Transactions on Medical Imaging, 32(11), pp 2127–39

[KRAM1991] Kramer, A. F. (1991). Physiological Metrics of Mental Workload: A Review of Recent Progress, Multiple-Task performance. In D. L. Damos (Eds.). London: Taylor and Francis. (1991) pp 279-328

[KARR2008]. Karray, F., Alemzadeh, M., Saleh, J.A. and Arab, M.N. (2008). Human-Computer Interaction: Overview on State of the Art, 1(1):137–159

[KELL2009] Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. Foundations and Trends® in Information Retrieval, 3(1—2), 1-224.

[KEMM2014] Kemmis, S., McTaggart, R., Nixon R. (2014). The Action Research Planner: Doing Critical Participatory Action Research, Springer, ISBN 978-981-4560-67-2

[KIER1988] Kieras, D. E. (1988). Towards a practical GOMS model methodology for user interface design. In The Handbook of Human-Computer Interaction. North-Holland, Amsterdam, pp 135–158

[KIER1993] Kieras, D. E. (1993). Using the Keystroke-Level Model to Estimate Execution Times. Unpublished Report, http://www.pitt.edu/~cmlewis/KSM.pdf, University of Michigan, accessed 31st March 2016.

[KIER1997] Kieras, D. E. (1997a). A Guide to GOMS model usability evaluation using NGOMSL. In The Handbook of Human-Computer Interaction, In M. Helander, T. Landauer, & P. Prabhu (Eds.) 2nd ed. North-Holland, Amsterdam, pp 733-766

[KIM2011] Kim, H., Cho, Y.S. and Do, E.Y. (2011). Using Pen-Based Computing in Technology for Health, Part IV, HCII 2011; 4, LNCS 6764, pp 192–201.

[KIM2015] Kim, H.S., Kim, H., Yoon, I., Jung, C.H. and Hwan, Y. (2015). User Performance Differences between Graphics Tablet and Mouse in Graphic Applications: Focus on Controllability and Accuracy. International Journal of Multimedia and Ubiquitous Engineering, 10(9), pp.167-174

[KIM2016] Kim, K. et al., 2016. Assisting people with visual impairments in aiming at a target on a large wall-mounted display. International Journal of Human Computer Studies, 86, pp s109–120.

[KIN2009] Kin, K., Agrawala, M., and DeRose, T. (2009). Determining the benefits of direct- touch, bimanual, and multifinger input on a multitouch workstation. In Proceedings of Graphics Interface, pages 119–124.

[KLEI2003] Klein, M., Sayama H, Faratin P, Bar-Yam Y. (2003). The Dynamics of Collaborative Design: Insights from Complex Systems and Negotiation Research, Concurrent Engineering, 11(3), pp 201–9

[KOCK2014] Kockelkorn, T. T. J. P., Schaefer-Prokop, C. M., Bozovic, G., Muñoz-Barrutia, A., van Rikxoort, E. M., Brown, M. S. and Van Ginneken, B. (2014). Interactive lung segmentation in abnormal human and animal chest CT scans. Medical Physics, 41, 081915

[KOTA2003] Kotani, K. and Horii, K. (2003). An analysis of muscular load and performance in using a pen–tablet system. J Physiol Anthropol 22, pp 89–95

[KOUT2007] Koutsabasis, P., Spyrou, T., Darzentas, J. (2007). Evaluating Usability Evaluation Methods: Criteria, Method and a Case Study. Human-Computer Interaction. Interaction Design and Usability Lecture Notes in Computer Science Volume 4550, 2007, pp 569-578

[KURZ2015] Kurzhals, K., Fisher, B., Burch, M. and Weiskopf, D. (2015). Eye tracking evaluation of visual analytics. Information Visualization (2015)

[KVAN2000] Kvan T. (2000). Collaboration Design: What is it? Automation in Construction, 9(4), pp 409-415

L

[LAMA] La Macchia, M., Fellin, F., Amichetti, M., Cianchetti, M., Gianolini, S., Paola, V., ... and Widesott, L. (2012). Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiation Oncology,7(1), 1.

[LANG1995] Lang, P.J. (1995). The Emotion Probe. American Psychologist, 50, pp 372 – 385

[LAW2014] Law, E.L. and Schaik, P. Van, (2014). Attitudes towards User Experience (UX) Measurement. International journal of Human computer studies, 72, pp 526–541

[LAZA2010] Lazar, J., Feng, J.H. and Hochheiser, H. (2010). Research Methods in Human-Computer Interaction. John Wiley & Sons, 2010.

[LEAP2016] Leap motion. (2016). Retrieved 1st September 2016 from https://www.leapmotion.com/.

[LEE2008] Lee, N., Smith, R.T., Laine, A.F. (2008). Interactive segmentation for geographicatrophy in retinal fundus images, Proc. 42nd Asilomar Conf. Signals, Systems and Computers, pp 655–658

[LEGA1998] Leganchuk, A., Zhai, S. and Buxton., W. (1998). Manual and cognitive benefits of two-handed input: an experimental study. ACM Trans. Comput.-Hum. Interact. 5, 4 (December 1998), pp 326-359

[LEON2007] Leong, J.J.H. et al., (2007). Visual search behaviour in skeletal radiographs: a cross-speciality study. *Clinical Radiology*, 62(11), pp 1069–1077

[LEWI1982] Lewis, C. H. (1982). Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report), IBM, RC-9265.

[LI2012] Li, J. Y., and Dang, J. W. (2012). Research and Improvement of Live-Wire Interactive Algorithm for Medical Image Segmentation. Applied Mechanics and Materials, 182-183, pp1065–1068

[LI2014] Li, X. et al., (2014). Remote heart rate measurement from face videos under realistic situations. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 4264–4271

[LILH2015] Lilholt, P.H., Jensen, M.H. and Hejlesen, O. (2015). Heuristic evaluation of a telehealth system from the Danish TeleCare North Trial. International Journal of Medical Informatics, vol 84, no. 5, pp 319-326

[LIN2013] Lin, C.J., Hsieh, T.L. and Lin, S.F. (2013). Development of staffing evaluation principle for advanced main control room and the effect on situation awareness and mental workload. J Nuclear Engineering and Design, 265, pp 137-144

[LONG2011] Longo, L. and Kane, B. (2011). A novel methodology for evaluating user interfaces in health care, 24th International Symposium on Computer-Based Medical Systems(CBMS), 27-30 June, pp1-6

[LOZA2011] Lozano, C., Jindrich, D. and Kahol, K. (2011). The impact on musculoskeletal system during multitouch tablet interactions. Proceeding CHI ’11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (August 2015), pp 825–828

[LUN1995] Lun, K.C. (1995). New user interfaces. International Journal of Bio Medical Computing, 39(1), pp 147–150

[LUND2011] Lundström, C., Rydell, T., Forsell, C., Persson, A. and Ynnerman, A. (2011). Multi-touch table system for medical visualization: application to orthopedic surgery planning. IEEE transactions on visualization and computer graphics; 17(12), pp 1775–84.

[LUPU2013] Lupu, R.G. and Ungureanu, F. (2013). A survey of eye tracking methods and applications. Bul Inst Polit Iasi, pp 71-86.

M

[MARS2000] Marshall, S. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. U.S. Patent 6,090,051, July 2000.

[MAZU2014] Mazur, L.M., Mosaly, P.R., Hoyle L.M., Jones E.L., Chera B.S., Marks L.B. (2014). Relating physician's workload with errors during radiation therapy planning. Practical radiation oncology, 4(2), pp 71-75

[MCWA2005] McWalter, E.J., Wirth W., Siebert M,, von Eisenhart-Rothe R.M., Hudelmaier M., Wilson D.R., Eckstein F. (2005). Use of novel interactive input devices for segmentation of articular cartilage from magnetic resonance images. Osteoarthritis and Cartilage, 13(1), pp 48-53

[MEDB2016] Radiation Therapy (2017), retrieved from http://www.medbroadcast.com/Procedure/GetProcedure/Radiation-Therapy on Feb.28, 2017.

[MEVI2016] Mevislab, retrieved on Sep.6, 2016 from http://www.mevislab.de/.

[MILL2000] Millen, D. R. (2000). Rapid Ethnography: Time Deepening Strategies for HCI Field Research. Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods and Techniques, pp 280-288

[MISH2008] Mishra, A., Wong, A., Zhang, W., Clausi, D. and Fieguth, P. (2008). Improved interactive medical image segmentation using enhanced intelligent scissors (eis). In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp 3083-3086

[MITK] The Medical Imaging Interaction Toolkit (MITK), http://www.mitk.org/, accessed on April.1st 2016.

[MIYA2001] Miyake, S. (2001). Multivariate Workload Evaluation Combining Physiological and Subjective Measures. International Journal of Psychophysiology, 40 (3), pp 233–238

[MOFF2006] Moffatt, S. et al., (2006). Using quantitative and qualitative data in health services research - what happens when mixed method findings conflict? [ISRCTN61522618]. BMC health services research, 6, pp 28

[MOLI2015]Molin, J., Lundström, C. and Fjeld, M. A. (2015). comparative study of input devices for digital slide navigation. J Pathol Inform 2015; 6:7

[MOLT2011] Moltz, J.H., Braunewell, S., Ruhaak, J., Heckel F., Barbieri S., Tautz L., Hahn H.K., Peitgen H.O. (2011). Analysis of variability in manual liver tumor delineation in CT scans, IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp 1974-1977

[MONA2008] Monahan, K., Lahteenmaki, M., McDonald, S and Cockton, G. (2008). An investigation into the use of field methods in the design and evalutaion of interactive systems. 22nd British Computer Society HCI group conference: Culture, Creativity, Interaction (HCI 2008), pp.99–108

[MORG2011] Morgan, J. F. and P. A. Hancock. (2011). The Effect of Prior Task Loading on Mental Workload An Example of Hysteresis in Driving. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53 (1), pp 75–86

[MOTT2013] Motti, L.G., Vigouroux, N., and Gorce. P. (2013). Interaction techniques for older adults using touchscreen devices : a literature review. 25`eme conf´erence francophone sur l'Interaction Homme-Machine, IHM'13, Nov 2013, Bordeaux, France

[MOSA2011] Mosaly, P.R., Mazur, L.M. (2011). Empirical Evaluation of Workload of the Radiation Oncology Physicist During Radiation Treatment Planning and Delivery, Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 55(1), pp 753-75

[MULT2011] Multi-Institutional Target Delineation in Oncology Group.(2011). Human–computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. Journal of digital imaging, 24(5), pp 794-803

N

[NIEL1990] Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces, Proc. ACM CHI'90 Conference (Seattle, WA, 1-5 April), 249-256. ics, 35(4), pp 379-390

[NIEL1994] Nielsen, J. and Levy, J. (1994). Measuring usability: preference vs. performance. Communications of the ACM, 37(4), pp 66–75

[NHS2016] Radiotherapy (2017), http://www.nhs.uk/Conditions/radiotherapy/Pages/introduction.aspx. Accessed on 20th June 2016

[NORO2013] Noronha, H., Sol, R. and Vourvopoulos, A. (2013). Comparing the Levels of Frustration between an Eye-Tracker and a Mouse: A Pilot Study. In Human Factors in Computing and Informatics, (Heidelberg, 2013), pp 107-121. Springer Berlin

[NUTT2000] Nutting, C., Dearnaley, D.P. and Webb,S. (2000). Intensity modulated radiation therapy: a clinical review. The British journal of radiology 73, no. 869, pp 459-469

O

[OLAB1997] Olabarriaga, S. and Smeulders, A. (1997). Setting the mind for intelligent interactive segmentation: Overview, requirements, and framework. In Proc. Int. Conf. Information Processing in Medical Imaging, pp 417–422

[OLAB2001] Olabarriaga, S. and Smeulders, A. (2001). Interaction in the segmentation of medical images: A survey. Med Image Anal 5(2), pp 127–142

[OYEW2011] Oyewole, S.A. and Haight, J.M. (2011). Determination of optimal paths to task goals using expert system based on GOMS model. Computer Human Behaviour, 27, pp 823–833

P

[PAPI1990] Papillo, J.F. and Shapiro, D. (1990). The cardiovascular system. In Principles of Psychophysiology: Physical, Social, and Inferential Elements, L.G. Tassinary (Ed.), pp 456-512 (Cambridge: Cambridge University Press).

[PARK2009]. Park, B. (2009). Physiology as a tool for HCI research: Promises and Pitfalls. J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2009, LNCS 5610, pp 141-148

[PARK2010] Park, Y. S., and Han, S. H. (2010). One-handed thumb interaction of mobile devices from the input accuracy perspective. IJIE 40, 6, pp 746-756

[PERA2011] Perandini, S., Faccioli, N., Inama, M. and Pozzi Mucelli, R. (2011). Freehand liver volumetry by using an electromagnetic pen tablet: accuracy, precision, and rapidity. Journal of digital imaging: the official journal of the Society for Computer Applications in Radiology, 24(2), pp 360-365

[POMP2003] Pomplun, M., and Sunkara, S. (2003). Pupil dilation as an indicator of cognitive workload in Human-Computer Interaction. In Proceedings of HCI International 2003: Vol 3, pp. 542-546

R

[RAMK2014] Ramkumar, A., Dolz, J., Kirisli, H.A., Varga, E., Schimek-Jasch, T. Nestle, U. Massoptier, L. Stappers, P.J., Niessen, W.J., Song, Y. (2014). Human-computer interaction in segmenting organs at risk for radiotherapy: a pilot study. Multimodal imaging towards individualized radiotherapy treatments, pp 69-79, ISBN: 978-94-6186-309-6

[RAMK2015] Ramkumar, A., Dolz, J., Kirisli, H.A., Varga, E., Adebahr S., Schimek-Jasch, T., Nestle, U., Massoptier, L., Stappers, P.J., Niessen, W.J. and Song, Y. (2016). User Interaction in Semi-Automatic Segmentation of Organs at Risk: a Case Study in Radiotherapy. J Digit Imaging, 29(2), pp 264-277

[RAMK2017] Ramkumar, A., Stappers, P.J., Niessen, W.J., Adebahr S., Schimek-Jasch, T., Nestle, U., and Song, Y. "Using GOMS and NASA-TLX to evaluate Human Computer Interaction process in interactive segmentation". International Journal of Human-Computer Interaction, Vol 33(2), 123-134, 2017

[RANI2002] Rani, P., Sims, J., Brackin, R. and Sarkar, N. (2002). Online Stress Detection using Psychophysiological Signal for Implicit Human– Robot Cooperation. Robotica, 20, pp 673 – 686

[RAUT2012] Rautaray, S.S. and Agrawal, A. (2012). Vision based hand gesture recognition for human computer interaction: a survey. Artificial Intelligence Review, 43(1), pp 1–54.

[ROGE2013] Rogers, M. L., Sockolow, P. S., Bowles, K. H., Hand, K. E., & George, J. (2013). Use of a human factor approach to uncover informatics needs of nurses in documentation of care. International Journal of Medical Informatics, 82(11), pp 1068–74

[ROTH2004] Rother, G., Kolmogorov, V. and Blake, A. (2004). “GrabCut” - Interactive foreground extraction using iterated graph cuts, in ACM Transactions on Graphics, Proceedings of ACM SIGGRAPH 2004, vol. 23, no. 3, pp 309–314

[ROSC1990] Roscoe, A. H. and G. A. Ellis. (1990). A Subjective Rating Scale for Assessing Pilot Workload in Flight: A decade of practical use, Technical report TR 90019, accessed on 1st April 2016

[ROSE1995] Rose, A., Shneiderman, B. and Plaisant, C. (1995). An applied ethnographic method for redesigning user interfaces. Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques, pp 115-122

[ROWE1998] Rowe, D.W., Sibert, J. and Irwin, D. (1998). Heart Rate Variability: Indicator of User State as an Aid to Human – Computer Interaction. In Conference on Human Factors in Computing Systems, pp 480-487

[RTOG2016] Guidelines for organs at risk delineation in thoracic radiation therapy.(2016). Based on RTOG guidelines. http://www.rtog.org/CoreLab/ContouringAtlases/LungAtlas.aspx, accessed on May.1, 2016

[RUBI2004] Rubio, S., Diaz, E., Martin, J. and Puente, J. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology, 53 (1), pp 61–86

[RUDD1996] Rudd, J., Stern, K. and Isensee, S. (1996). Low vs. high-fidelity prototyping debate. interactions, 3(1), pp 76-85

S

[SADE2009] Sadeghi, G.H.H.M., Tien, G. and Atkins, M.S. (2009). Hands-free interactive image segmentation using eyegaze. SPIE Medical Imaging conference.

[SAIT2010] Saitwal, H., Feng, X. and Walji, M. (2010). Assessing performance of an electronic health record (EHR) using cognitive task analysis, Int. J. Medical Informatics, 79 (7), pp 501-506

[SCHR2006] Schrepp, M. and Fischer, P.A. (2006). GOMS model for keyboard navigation in web pages and web applications**.** Lecture Notes in Computer Science, 4061, pp 287-294

[SCHU2003] Schultz, M., Gill, J., Zubairi, S., Huber, R., Gordin, F. (2003) Bacterial contamination of computer keyboards in a teaching hospital. Infect Control Hosp Epidemiol 4(24), pp 302-303

[SHAR2014] Sharma, C. and Dubey, S.K. (2014). Analysis of eye tracking techniques in usability and HCI perspective. 2014 International Conference on Computing for Sustainable Global Development, INDIACom 2014, pp 607-612

[SHER2005] Sherbondy, A.J., Holmlund, D., Rubin, G.D., Schraedley, P.K., Winograd, T. and Napel, S. (2005). Alternative input devices for efficient navigation of large CT angiography data sets. Radiology 234, pp 391-398

[SHI2007] Shi, Y. et al., (2007). Galvanic Skin Response (GSR) as an Index of Cognitive Load. CHI EA ’07 CHI '07 Extended Abstracts on Human Factors in Computing Systems, pp 2651-2656

[SHIF1991] Shiffman, S., Wu, A. W., Poon, A.D., Lane, C.D., Middleton, B., Miller, R.A, Masarie, F. E., Cooper, G.F., Shortliffe, E.H. and Fagan, L.M. (1991). Building a speech interface to a medical diagnostic system, 1991, IEEE Expert, pp 41-50

[SIMS2009] Sims, R., Isambert, A., Grégoire, V., Bidault, F., Fresco, L., Sage, J., Mills, J., Bourhis, J., Lefkopoulos, D., Commowick, O., Benkebil, M., Malandain, G. (2009). A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. Radiother Oncol 93(3), pp 474–478.

[SLOT2003] Slotman, B.J., Leer, J.W.H. (2003). Infrastructure of radiotherapy in the Netherlands: evaluation of prognoses and introduction of a new model for determining the needs, Radiotherapy and Oncology, 66, pp 345–349

[SMEL1995] Smelcer, J.B. (1995). User errors in database query composition. International Journal of Human–Computer Studies, 42, pp 353–381

[SOWE2009] Sowell, R., Liu, L., Ju, T., Grimm, C., Abraham, C., Gokhroo, G., and Low, D. Volume Viewer. (2009). An Interactive Tool for Fitting Surfaces to Volume Data, EUROGRAPHICS Symposium on sketch-Based interfaces and Modeling

[STEG2012] Steger, S., and Sakas, G. (2012). FIST: Fast Interactive Segmentation of Tumors. H. Yoshida, G. Sakas, & M. G. Linguraru, Eds.), Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and

Lecture Notes in Bioinformatics), Berlin, Heidelberg: Springer Berlin Heidelberg, Vol. 7029, pp 125–132

[STER2001] Stern, R.M., Ray, W.J. and Quigley, K.S. (2001). Psychophysiological Recording (New York: Oxford University Press).

[STOR2011] Story, M.F. (2011). FDA Human Factors Draft Guidance Document, Retrieved on May. 3, 2016 from www.fda.gov/downloads/MedicalDevices/ Device Regulation and Guidance/ Human Factors/UCM320901.pdf, 2011,

[STRI2016] Stribling, J.C. and Richardson, J.E. (2016). Placing wireless tablets in clinical settings for patient education. Journal of the Medical Library Association: JMLA, 104(2), pp 159–64

[SUN2013] Sun, S., Sonka, M., and Beichel, R. R. (2013). Lung segmentation refinement based on optimal surface finding utilizing a hybrid desktop/virtual reality user interface. Computerized Medical Imaging and Graphics, 37(1), pp 15–27

[SZÉK2013] Székely, A., Talanow, R., and Bágyi, P. (2013). Smartphones, tablets and mobile applications for radiology. European Journal of Radiology, 82(5), pp 829–836

T

[TAVE2015] Taveira, A.D. and Choi, S.D. (2015). Computer Input Devices and the Elderly: A Musculoskeletal Disorder Focus. J. Zhou and G. Salvendy (Eds.): ITAP 2015, Part II, LNCS 9194, pp 333–340

[TERW2015] Trewin, S., Marques, D. and Guerreiro, T. (2015). Usage of Subjective Scales in Accessibility Research. ASSETS ’15 The 17th International ACM SIGACCESS Conference on Computers and Accessibility, (October), pp 59-67

[TOBI2016] Tobii. (2016). http://www.tobii.com/, Retrived 1st september 2016

U

[UCLA2016] Cervical, Endometrial, Vaginal and Vulvar Cancers, Gynecologic Brachytherapy Treatment. http://radonc.ucla.edu/gynecologic-brachytherapy-treatment, Retrived 1st september 2016

V

[VARG2013] Varga, E., Pattynama, P. M. T., and Freudenthal, A. (2013). Manipulation of mental models of anatomy in interventional radiology and its consequences for design of human–computer interaction. Cognition, Technology & Work, 15(4): pp 457-473

[VELA2013] Velazquez, E. R., Aerts, H. J. W., Yuhua, G., Goldgof, D. B., Ruysscher, D., Dekker, A., Korn, R., Gillies, R. J., and Lambin, P. A. (2013). semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen, 105(2), pp 167-173

[VERM2007] Vermeeren, A., Bekker,M.M., van Kesteren, I.E.H. and De Ridder, H. (2007). Published by the British Computer Society People and Computers XXI – HCI… but not as we know it: Proceedings of HCI 2007

W

[WACH2008] Wachs, J.P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C. and Smith, M. A. (2008). Gesture-based Tool for Sterile Browsing of Radiology Images, Journal of the American Medical Informatics Association, 15 (3), pp 321-23

[WANG2012] Wang, Y., Dang, J., Du, X. and Yang, Y. (2012). Study on the interactive serial medical images segmentation. In International Conference on Automatic Control and Artificial Intelligence (ACAI 2012) pp 1988-1992

[WARD2002] Ward, R.D., Marsden, P.H., Cahill, B and Johnson, C. (2002). Physiological Responses to Well-Designed and Poorly Designed Inter- faces, In CHI 2002 Workshop on Physiological Computing (Minneapolis, MN, USA)

[WARD2003] Ward, R.D and Marsden, P.H. (2003). Physiological responses to different WEB page designs. International Journal of Human – Computer Studies, 59, pp 199 – 212

[WEST2007] West, R.L. and Nagy, G. (2007). Using GOMS for modelling Routine Tasks Within Complex Sociotechnical Systems: connection Macrocognitive Models to Microcognition. Journal of Cognitive Engineering and Decision Making. June 20, vol. 1,no. 2, pp 186-211

[WHAR1994] Wharton, C., Rieman, J., Lewis, C. and Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In Usability inspection methods, pp 105-140, John Wiley & Sons, Inc..

[WHIT2013] Whitfield, G.A., Price, P., Price, G.J. and Moore, C.J. (2013). Automated delineation of radiotherapy volumes: are we going in the right direction? Br J Radiol 86(1021):20110718.

[WHO2017] WHO, All Cancers (excluding non-melanoma skin cancer), Estimated Incidence, Mortality and Prevalence Worldwide in 2012, Retrieved on Sep.3 2016, from http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

[WILS2000] Wilson, G.M and Sasse, M.A. (2000). Do Users Always Know What's Good For Them? Utilizing Physiological Responses to Assess Media Quality. In HCI 2000: People and Computers XIV – Usability or Else, pp 327 - 339

[WILS2001] Wilson, G.M. (2001). Psychophysiological Indicators of the Impact of Media Quality on Users. In CHI 2001 Doctoral Consortium (Seattle, WA, USA: ACM Press), pp 95- 96

[WINT1984] Winton, W., Putnam, L and Krauss, R. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. Journal of Experimental Social Psychology, 20, pp 195 – 216

[WU2010] Wu, K., Ung, Y.C., Hwang, D., Tsao, M.S., Darling, G., Maziak, D.E., Tirona, R., Mah, K., Wong, C.S. (2010). Auto-contouring and manual contouring: which is the better method for target delineation using 18F-FDG PET/CT in non-small cell lung cancer? J Nucl Med 51(10) pp1517–23

X

[XIAN2010] Xiang, L.X.L. and Xiaoli, C.X.C. (2010). The research on performance of automobile human-machine interface based on BHR-GOMS behavior model. In: Intelligent computer intelligent systems (ICIS), IEEE international Conference, 2, pp 174-178

Y

[YANG2010] Yang, W., Cai, J., Member, S., Zheng, J. and Luo, J. (2010). User-friendly Interactive Image Segmentation through Unified Combinatorial User Inputs(c), pp 1–10

[YEN2009] Yen, P.Y. and Bakken, S. (2009). A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol – An Example from a Web-based Communication Tool for Nurse Scheduling. AMIA annual symposium proceedings, pp 714-718.

[YEO2011] Yeo, S.Y., Xie, X., Sazonov, I. and Nithiarasu, P. (2011). Geometrically induced force interaction for three-dimensional deformable models. IEEE Trans Image Process 20(5), pp 1373-1387

[YURK2010] Yurko, Y.Y., Scerbo, M.W., Prabhu, A.S., Acker, C.E. and Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simulation in Healthcare: Journal of the Society for Simulation in Healthcare, 5(5), pp 267-71

Z

[ZEWE2014] Zewei, Z., Tianyue, W., Li, G., Tingting, W., & Lu, X. (2014). An interactive method based on the Live Wire for segmentation of the breast in mammography images. Computational and mathematical methods in medicine, 2014.

[ZHAN2003] Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L. and Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. Journal of Biomedical Informatics, 36(1–2),pp 23–30.

[ZHAO2013] Zhao, F., and Xie, X. (2013). An Overview of Interactive Medical Image Segmentation, Annals of BMVA, 7, pp 1–22.

[ZHOU2013] Zhou, W., and Xie, Y. (2013). Interactive medical image segmentation using snake and multiscale curve editing. Computational and Mathematical Methods in Medicine, Article ID 325903

[ZHU2009] Zhu, Y., Shen, T., Lopresti, D. and Huang, X. (2009). Interactive polygons in region-based deformable contours for medical images. IEEE International Symposium on Biomedical Imaging: From Nano to Macro pp 37–40

Acknowledgements

I, Anjana Ramkumar have been working at the faculty of Industrial Design engineering, Delft university of Technology since November 2012 till November2016. In this section, I would like to take this opportunity to express my heartfelt thanks to my supervisors and colleagues, for generously sharing their knowledge and experiences with me, as well as my family and friends for their unconditional love and support.

First of all, I would like to thank my promoters Prof. Dr. W.J. Niessen and Prof. Dr. P.J. Stappers and co-promoters Dr. Y. Song and Dr.ir. A. Freudenthal for their guidance and support. I enjoyed working with both my promoters as both of them had two different backgrounds. My constant discussion with both my promoters helped me to gain knowledge about image segmentation and in the field of design. Also, the monthly meetings with you were very helpful as it always helped me to think more. You both are always my inspiration and I look forward to working with you in near future, if I get an opportunity. Dear Adinda. You brought me to the world of design and taught me what user research is about. You are no more with us now but you will be remembered always. Dear Wolf, I admire your thoroughness, dedication, hardworking nature and loyalty. Our long discussions during the weekly meetings and other times always motivated and energized me to explore a lot of new things. Mainly because of you, my knowledge in computational algorithms and statistics have improved. I look forward to work with you more in the future. Thank you for everything. My mentor Dr. S.U.Boess, you guided me throughout the graduate school programs and advised me to take different courses during each step of my PhD research. You attend all my meetings and used to send all the meeting noted promptly. Thank you.

The presented research is part of Software for the Use of Multi-Modality images in External Radiotherapy (SUMMER) project which is funded by European Commission (FP7-PEOPLE-2011-ITN) under grant agreement PITN-GA-2011-290148. I would like to thank the funding agency for providing financial support during the full journey of PhD. I would like to express my thanks and appreciations to my SUMMER project consortium project leaders: Mr.Laurent, Dr.Ursula, Dr.Anne, Dr.Umberto, Dr.Katja and Dr.Wolfgang, for their valuable advice during my research period. I would like to mention my special thanks to Aquilab for helping me with my prototype development. Also, I would like to thank doctors from University Medical Centre, Freiburg, Germany and Institut Claudius Regaud, Oncopole, Tulouse, France for giving their valuable time to participate in my study. I had a very great time with the 12 other SUMMER researchers and have learnt a lot from them in both professional and personal things.

My PhD committee members: Prof.Dr. T. Brunner, Dr. S. D. Olabarriaga, Prof.dr.ir. J. M. P. Geraedts, Prof.dr.ir. R.H.M Goossens, Rector Magnificus, Prof.dr. W.J. Niessen, Prof.dr. P.J.Stappers, and Dr. Y. Song, thank you for your time and efforts to read my thesis, giving the feedback, as well as willingness to attend my defence.

I would like to thank the Radiotherapy department at the Erasmus medical centre, Rotterdam who took the time to meet me and discuss my work and allowing me to do some initial observational studies. I would like to Dr. Jiehui Jiang for collaborating and contributing for my PhD. Thanks for giving me the opportunity to visit China and arranging participants for my study. I would thank Ms. Qing Ding and Mr. Dongxing Gu who helped while conducting the experiment in China. I would also like to thank the radiotherapy department of Renji hospital and nuclear medicine department of the Huashan hospital for providing participants for our study.

I would like to thank my faculty, department and office colleagues, with whom i was very fortunate to work with and have learnt a lot of things from them. I would like to thank my office mates Anet, Bahar, and Prang with whom I shared a common office during my PhD. They were kind enough to listen to my problems during my tough time and stood by me during my ups and downs, in this PhD journey. I would also like to thank my former colleagues Cecile and Edit, who shared their knowledge with me and helped me whenever i had some question. Edit you were an extraordinary mentor as well as good friend. Thank you for everything.

Life outside research was fun with my friends. I have been fortunate to meet several extraordinary people in my life. Their friendship was a strong umbrella which helped me through several storms. Especially my Indian friends who were with me during my highs and lows. I will be taking back a lot of memories (friday dinners, potluck lunch and dinner, exploring new restaurants, weekend outings, playing badminton, mafia's, truth or dare, festival celebrations etc..) which i had with them. My thanks to Abhigyan, Amy, Arun M, Arun J, Arun, Bharadwaj, Freddy, Gautham, J.K, Kamakshi, Laksh (Lakshmi), Lakshmi, Mohit, Nandini, Niru, Nishanth, Noopur, Paru, Prasanth, RAP (R.A.Prasanth), Raghu, Ranju, Shiva, Sriram, Supraja, Seshanji, Sruthi, Syam, Vasu, Venky, Viknesan, Vinay. I would also like to thank all other Indian and Dutch friends in Netherlands whom i met during this four years.

Special thanks to my friends in UK, US (Anu and poopal) and India for all their support.

Finally, I owe my gratitude to my grandparents, parents, in-law's, my sister, my brother-in-law's, and other family members who lent me constant support. Last but not the least, my husband Dr. Vijay Harish, who was a constant source of encouragement, my backbone, my inspiration to complete this work and carrying me through difficult situations. Even though we were in a long distant relationship, you stood by me all the times whenever I need your support. I am blessed to be part of such a family who has lot of faith in my abilities and supported me in all my aspirations. I hope this work makes them proud.

Author's Publications

Journals

1. **Anjana Ramkumar**, Jose Dolz, Hortense A.Kirisli, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle, Laurent Massoptier, Edit Varga, Pieter Jan Stappers, Wiro J.Niessen and Yu Song. "User Interaction in Semi-Automatic Segmentation of Organs at Risk: a Case Study in Radiotherapy". Journal of Digital Imaging, 29(2), 264-277, 2016.
2. **Anjana Ramkumar**, Pieter Jan Stappers, Wiro J.Niessen, Sonja Adebahr, Tanja Schimek-Jasch, Ursula Nestle and Yu Song. "Using GOMS and NASA-TLX to evaluate Human Computer Interaction process in interactive segmentation". International Journal of Human-Computer Interaction, Vol 33(2), 123-134, 2017 .
3. Yu Song, Jesse Hoeksema, **Anjana Ramkumar**, Johan Molenbroek. "A landmark based 3D parametric foot model for footwear customization". Accepted in the International Journal of the Digital Human, January 2017.

Conferences

4. **Anjana Ramkumar**, Yu Song ,Wiro J. Niessen , Pieter Jan Stappers. "Design issues of the existing medical image segmentation software : A case study using observational, heuristic and think-aloud methods". Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare June 2016, vol 5 no 1, pp 1-8.
5. **Anjana Ramkumar,** Yu Song et al., "Comparison of Heuristic Evaluation and Think Aloud Methods A Study in Radiotherapy Contouring Software". Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare 2014 June; vol 3(1): 230-237.
6. **Anjana Ramkumar**, Jose Dolz, Hortense A Kirisli, Tanja Schimek-Jasch, Sonja Adebahr, Ursula Nestle, Laurent Massoptier, Edit Varga, Pieter Jan Stappers, Wiro J Niessen, Yu Song. "Human computer interaction in segmenting organs at risk for radiotherapy: A pilot study". Multimodal imaging towards individualized radiotherapy treatments, Editors: Massoptier and Song, pp. 69-79, July 2014.
7. **Anjana Ramkumar**, Edit Varga, Anne Laprie, Wiro J Niessen, Adinda Freudenthal . "A pilot study of Pen-Tablet interaction in radiotherapy contouring using orthogonal and non-orthogonal views". Design of Medical Devices Conf – Europe (DMD'13 – Delft), October 2013.

8. **Anjana Ramkumar**, Edit Varga, Wiro J Niessen, Adinda Freundethal. "Exploring input devices for contouring in external radiotherapy". Innovative imaging to improve radiotherapy treatments, Lulu Enterprises Inc Ed, 1:29-34, July 2013.

9. Anet Aselmaa, Richard HM Goossens, Anne Laprie, **Anjana Ramkumar**, Soléakhéna Ken, Adinda Freudenthal." External radiotherapy treatment planning–situation today and perspectives for tomorrow". Innovative imaging to improve radiotherapy treatments, Lulu Enterprises Inc Ed, 1:91-98, July 2013.

10. Adinda Freudenthal, Anet Aselmaa, **Anjana Ramkumar**, Anne Laprie, . Laurent Massoptier." Human Factors Engineering for Radiotherapy planning software". Innovative imaging to improve radiotherapy treatments". Innovative imaging to improve radiotherapy treatments, Lulu Enterprises Inc Ed, 1:129-134, July 2013.

Featured Article

11. **Anjana Ramkumar.** "Designing interaction for radiotherapy planning". Communicating Design, Form magazine, Sep/Oct 2016 issue.

Project deliverables

12. **Anjana Ramkumar**, Anet Aselmaa, Yu Song." User interface concept and new workflow report". http://summer-project.eu/wp-content/uploads/2014/09/D2.2_UI-concept_report_WEB.pdf, 2014.

13. Anet Aselmaa, Richard HM Goossens, Anne Laprie, Soléakhéna Ken, Tobias Fechter, **Anjana Ramkumar**, Adinda Freudenthal. "Workflow analysis report". http://summer-project.eu/wp-content/uploads/2013/10/D2.1_Workflow_analysis_report_WEB.pdf, 2013.

14. **Delft University of Technology**. "Usability feedback report". http://summer-project.eu/wp-content/uploads/2016/01/D2.3_Usability_Feedback_v03_1.pdf**,** 2015

Ms. Anjana Ramkumar was born on 16th October 1989 at Renukoot, Uttar Pradesh, India. She had her schooling from the state of Tamil Nadu located in southern India. Ms. Ramkumar holds a Bachelors degree in Medical Radiological Technology from the Amrita Institute of Medical Sciences and Research Centre, Kochi, India. The title of her Bachelor graduation project was Assessment of Gross Tumour Volume by Manual Vs Auto Contouring using PET-VCAR .
After attaining her Bachelor degree, Ms. Ramkumar moved to the United Kingdom for her Masters in Medical Physics at the University of Surrey. Her Master thesis was on the topic Commissioning 4D CT for treatment of stage 1 lung cancer which was a collaboration between the Royal Sussex cancer centre, Brighton, and the University of Surrey. During her undergraduate and postgraduate hospital training, she was also introduced to practical work in Radiotherapy Treatment Planning for 3 years.
Since 2012 November, Ramkumar started her doctoral studies at Delft university of Technology, The Netherlands and was working for the SUMMER project as a Marie Curie research fellow, funded by European Commission (FP7-PEOPLE-2011-ITN) under grant agreement PITN-GA-2011-290148. After submitting the thesis in November 2016, Ms. Ramkumar joined Amrita Institute of Medical Sciences and Research Centre, Cochin, Kerala, India as a Post-Doctoral fellow and a research manager.