Incorporating sensemaking perspective in design
Supporting physicians during the contouring tasks in radiotherapy

Aselmaa, Anet

DOI
10.4233/uuid:a9ff69cd-e5a5-451e-a6c8-0681b633927d

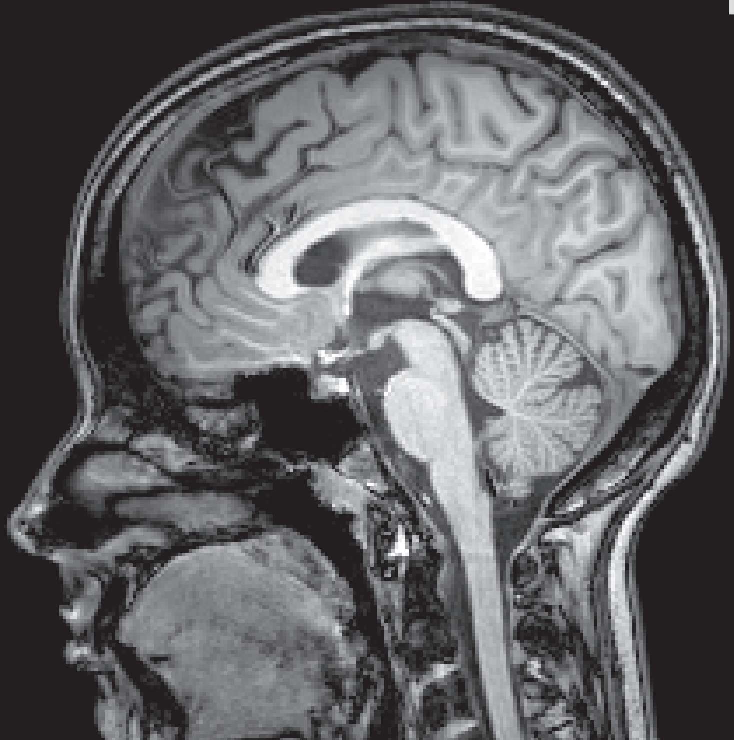
Publication date
2017

Document Version
Final published version

Citation (APA)
Aselmaa, A. (2017). *Incorporating sensemaking perspective in design: Supporting physicians during the contouring tasks in radiotherapy*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a9ff69cd-e5a5-451e-a6c8-0681b633927d

Incorporating sensemaking perspective in design

Supporting physicians during the contouring tasks in radiotherapy



Anet ASELMAA

Incorporating sensemaking perspective in design

Supporting physicians during the contouring tasks in 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
woensdag 20 december 2017 om 10:00 uur

door

Anet ASELMAA
Master of Science, Tallinn University of Technology
geboren te Võru, Estland

This dissertation has been approved by the

Promotor: Prof.dr. ir. R.H.M. Goossens

Copromotor: Dr. Y. Song

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr.ir. R.H.M. Goossens	Delft University of Technology
Dr. Y. Song	Delft University of Technology

Independent members:

Prof.dr. MD A. Laprie	Institut Universitaire du Cancer de Toulouse-Oncopole, France
Prof.dr. MD U. Nestle	Kliniken Maria Hilf, Mönchengladbach, Germany
Prof.dr.ir. J.M.P. Geraedts	Delft University of Technology
Prof.dr.ir. J.H. Eggen	Eindhoven University of Technology
Prof.dr. P.J. Stappers	Delft University of Technology

Cover design: A. Aselmaa
ISBN/EAN: 978-94-028-0890-2
Printed at: Ipskamp Printing

The research has received funding from the European Union Seventh Framework Programme (FP7-PEOPLE-2011-ITN) under grant agreement PITN-GA-2011-290148

TABLE OF CONTENTS

SUMMARY

Radiotherapy is a type of cancer treatment that uses high energy radiation to shrink tumors by destroying cancer cells. It is estimated that 52 per cent of cancer patients can potentially benefit from this type of treatment (Delaney et al. 2005). Planning radiotherapy treatment is a complicated multi-disciplinary process (Aselmaa et al. 2013b). One of the most critical and cognitively challenging steps in the workflow for planning treatment is contouring. Through a complicated underlying cognitive process of ‘sensemaking’, physicians draw the visible boundary of the tumor (i.e. gross tumor volume, GTV) and surrounding organs that are also at risk, as identified in medical images, based on the synthesis of different types of data as well as their knowledge and experience.

The accelerated development of technology and the increasing amount of data and clinical knowledge pose new challenges for designers of software systems to support physicians’ cognitive processes. In the early phase of designing these systems, in addition to having an explicit understanding of users, tasks and contexts, it is important for designers to incorporate a theoretical understanding of cognitive processes, such as sensemaking, into the design process. Based on case studies conducted in the context of radiotherapy, the aim of this thesis is:

> **To provide the means to understand physicians’ sensemaking process during the early phase of design in order to design software that is well-fitted to the clinical workflow.**

This research project was exploratory in nature and touched many fields, such as radiotherapy, computational algorithms, theoretical and practical aspects of cognition, and design. It was conducted in collaboration with oncologists, medical physicists, computer scientists and industry partners.

The usual starting point for research and design is to explore the context. In order to understand the sensemaking processes in the radiotherapy context in relation to the overall workflow of treatment planning and other cognitive processes, we took a broad view for conducting our analysis. Chapter 2 sets out our workflow analysis of radiotherapy, focusing on the treatment planning process and the task of contouring. This analysis served two purposes: (1) to generate detailed knowledge regarding the context, and (2) to facilitate communication among the members of the multi-disciplinary research team. We identified many areas for improvement, including the need for better

comprehension of physicians’ cognitive processes when contouring in order to provide better support to design clinically well-fitting software.

Chapter 3 describes our further analysis of the cognitive processes involved in the contouring task through ethnographic research, a literature review and interviews. We identified that the main cognitive challenge for physicians is making sense of existing data, and more specifically comprehending the relevant parts of the entire body of data. In addition, we identified 29 medical factors that influence physicians’ reasoning during the contouring task and divided them into three categories: treatment context, tumor context and tumorous areas.

In order to arrive at a holistic view for the design of software to support sensemaking, technology aspects also needed to be considered. Computational algorithms can facilitate the contouring task with the incorporation of semi-automatic or fully automatic contouring methods. Fully automated contouring has known limited success (Bauer et al. 2013), but some semi-automatic methods are being increasingly embedded into commercial software solutions (Sykes 2014). Numerous different methods to this end have been researched and proposed (Olabarriaga and Smeulders 2001); however, they are often tailored for specific contouring tasks and types of tumors. The challenge for the designer is to comprehend the computational possibilities and limitations for maximizing the combined power of software and the human brain. Chapter 4 presents an analysis of automated contouring methods based on an object-oriented modeling approach. We identified the main design challenges for incorporating different automated contouring methods into radiotherapy planning software and divided them into four categories: general usability, navigation, workflow, and flexibility of interactions.

Based on the findings in Chapters 2 and 3 and Section 4.1, we describe the design and implementation of research software prototypes in Section 4.2. We then conducted studies on the GTV contouring task, which are described in Chapters 5 and 6.

Successful software is aligned with the clinical workflow. However, by adjusting the software to fit with the workflow, one may influence the way physicians perform the task and consequently the sensemaking process and the results. In Section 5.1, we describe an experiment to explore the impact of workflow-specific contouring software design on eight physicians. We identified that sub-region based workflow of tumor contouring minimized the interactions with different medical imaging datasets and reduced the cognitive demands required for information fusion. At the same time, this alternative task workflow resulted in an average decrease of 16 per cent of the resulting GTV contours. Therefore, designers should be cautious when designing software that is closely aligned with the workflow as this can influence the sensemaking process and consequently the clinical outcomes. Based on this study, we concluded that the next steps in our research should follow the same GTV contouring workflow as in current clinical practice.

In Section 5.2, we propose a two-step approach to support designers to gain insight into the sensemaking process and translate this into design requirements. This approach consists of: (1) modelling sensemaking in context, and (2) in-depth analysis of software interaction (patterns) in relation to that model. We applied this approach to a study of the tumor contouring task using the software prototype involving eight physicians. We captured the interactions with the software prototypes into log files, which we then visualized as timelines (Section 5.3) and used as a basis for inferring sensemaking activities.

We identified three main phases in the contouring task: familiarization, action and evaluation. We then identified interaction patterns within each phase. For instance, long navigation (going through a wide range of 2D image slices) occurred primarily during the familiarization and evaluation phase. We then linked the interaction patterns with sensemaking activities. For example, during the familiarization phase, the main sensemaking activities were building initial frames and exploratory information seeking, while during the evaluation phase, the physicians were primarily occupied with focused information seeking. Based on these findings, we identified five main areas in which to improve support of sensemaking processes: (1) to enable effective initial frame development; (2) to support intuitive navigation within and between datasets; (3) to support detecting regions of interest; (4) to enable additional methods for contour evaluation; and (5) to improve general efficiency by reducing time and physical effort required. We concluded that our proposed two-step approach proved beneficial for gaining detailed insight into the sensemaking process, and for deriving design requirements for increasing sensemaking support.

We confirmed that increasing time efficiency is an important challenge when designing sensemaking support software. We had already explored the possibility of incorporating more technical automation in Chapter 4; however, as regards using automation tools in sensemaking support software solutions, it was unclear how automation might influence the reasoning process and/or the contouring result. Therefore, we conducted a further study to analyze the influence of a common automated contouring tool known as between slices contour interpolation. We identified that, using this tool, physicians were able to perform the task more efficiently because they did not have to draw the initial contour. Furthermore, we observed that the results obtained via interpolation reached higher consistency among physicians. At the same time, this automation also influenced the contouring process and meant that physicians spent less time examining the data and the results.

The contributions of the research into designing for sensemaking conducted in this thesis are two-fold: we developed more knowledge regarding physicians' sensemaking process during the contouring task, and proposed a new approach for designing sensemaking support software. This can be viewed as a first step towards software that effectively supports physicians' sensemaking during tumor contouring. In addition, our proposed approach of incorporating sensemaking perspectives should assist designers tackling similar design challenges.

SAMENVATTING

Radiotherapie is een vorm van kankerbehandeling waarbij hoogenergetische straling wordt gebruikt om kankercellen te vernietigen en zo tumoren te laten krimpen. Naar schatting 52% van de kankerpatiënten zou kunnen profiteren van dit type behandeling (Delaney et al. 2005). De planning van radiotherapie is een complex, multidisciplinair proces (Anet Aselmaa, Goossens, Laprie, Ken et al. 2013). Een van de stappen in de planning van de behandeling is contourering, een kritieke stap die veel cognitieve inspanning vereist. Hierbij stelt de arts op de medische beelden de visuele randen vast van de tumor (het 'gross tumor volume' of GTV) en de omringende organen die risico lopen. Dit wordt gedaan door middel van een complex achterliggend cognitief proces van 'sensemaking' op basis van een synthese van zowel verschillende soorten data als kennis en ervaring.

De zich steeds sneller ontwikkelende technologie en de groeiende hoeveelheid data en klinische kennis leveren voor ontwerpers nieuwe uitdagingen op bij het ondersteunen van de cognitieve processen van artsen door middel van softwaresystemen. Tijdens de vroege fases van het ontwerp is het – ook met expliciet inzicht in de gebruikers, taken en contexten – niet eenvoudig om theoretisch inzicht in cognitieve processen zoals sensemaking mee te nemen in het ontwerpproces. Daarom beoogt dit proefschrift om op basis van case studies in de context van radiotherapie:

> **Tijdens de vroege ontwerpfase middelen te bieden om inzicht te krijgen in het sensemaking-proces van artsen om hiermee klinisch goed bruikbare software te ontwikkelen.**

Het onderzoek was verkennend van aard en had betrekking op een groot aantal vakgebieden, waaronder radiotherapie, rekenalgoritmes, theoretische en praktische aspecten van cognitie en het ontwerp. Daarom was de promovenda lid van een multidisciplinair team en is zij ondersteund door oncologen, medisch natuurkundigen, informatici en partners uit het bedrijfsleven.

Eén van de uitgangspunten voor onderzoek in ontwerpen is om onderzoek te doen naar de context. Om inzicht te krijgen in het sensemaking-proces in de context van radiotherapie in relatie tot de algemene workflow (werkstroom) en andere cognitieve processen is voor de analyse een breed perspectief gekozen. In hoofdstuk 2 wordt een analyse gepresenteerd van de workflow bij radiotherapie, met speciale nadruk op het planningsproces voor de behandeling en het maken van de contour rondom de tumor. Het doel van deze analyse is tweeledig: (1) het genereren van gedetailleerde kennis over de context, en (2) het faciliteren van communicatie tussen de leden van het multidisciplinaire onderzoeksteam. Er zijn veel punten voor verbetering gevonden, waaronder de noodzaak van het verkrijgen van meer inzicht in de cognitie van de arts tijdens het maken van de contour, zodat betere ondersteuning kan worden geboden door klinisch goed passende software.

Verdere analyse van de cognitieve processen die een rol spelen bij het maken van de contour is gedaan door middel van etnografisch onderzoek, literatuuronderzoek en interviews (hoofdstuk 3). Hierbij is vastgesteld dat de grootste cognitieve uitdaging voor de arts bestaat uit het interpreteren van de bestaande data, met name om inzicht te krijgen in de relevante onderdelen van alle data. Daarnaast zijn er 29 factoren geïdentificeerd die de gedachtegang van de arts tijdens het maken van de contour beïnvloeden, verdeeld over drie categorieën: de context van de behandeling, de context van de tumor en de gebieden waar zich tumoren bevinden.

Om een holistisch overzicht te krijgen voor het ontwerp van software die de sensemaking ondersteunt, moeten ook technologische aspecten worden meegenomen. Het maken van de contour kan worden ondersteund met rekenalgoritmes als er een deels of geheel geautomatiseerde methode wordt gebruikt. Geheel geautomatiseerd contour generatie is in beperkte mate succesvol gebleken (Bauer et al. 2013), terwijl bepaalde gedeeltelijk geautomatiseerde methoden steeds vaker onderdeel zijn van commerciële softwareoplossingen (Sykes 2014). Tegelijkertijd zijn er talloze verschillende methoden onderzocht en voorgesteld (Olabarriaga en Smeulders 2001), hoewel deze vaak specifiek bedoeld zijn voor bepaalde contoureringstaken en soorten tumoren. Voor de ontwerper is het de uitdaging om zowel de mogelijkheden als beperkingen met betrekking tot het rekenvermogen te begrijpen en de gecombineerde kracht van computers en menselijke intelligentie te maximaliseren. In hoofdstuk 4 wordt een analyse gepresenteerd van geautomatiseerde contourering op basis van objectgeoriënteerde modellering. De voornaamste uitdagingen met betrekking tot het inbouwen van verschillende methoden voor geautomatiseerde contourering in planningsoftware voor radiotherapie zijn geïdentificeerd en verdeeld over vier categorieën: algemene bruikbaarheid, navigatie, workflow en de flexibiliteit van de interacties.

Op basis van de bevindingen uit hoofdstuk 2 en 3 en paragraaf 4.1 worden in paragraaf 4.2 het ontwerp en de implementatie van de prototypes van de onderzoekssoftware gepresenteerd op basis van dit prototype. Onderzoek naar het contoureren van het GTV wordt gepresenteerd in hoofdstuk 5 en 6.

Een goed ontwerp zal meestal aansluiten op de klinische workflow. Maar door de software aan te passen aan de workflow kan de manier waarop de arts de taak uitvoert worden beïnvloed, met gevolgen voor het sensemaking-proces en de uitkomst. In paragraaf 5.1 wordt een experiment gepresenteerd waarmee de invloed van een werkstroomspecifiek ontwerp van contoureersoftware op acht artsen is onderzocht. Het bleek dat een workflow voor de contourering van tumoren op basis van subgebieden de interacties met verschillende medische visuele datasets minimaliseerde en de cognitieve inspanning van de artsen voor het combineren van informatie verminderde. Tegelijkertijd resulteerde de alternatieve workflow voor de taak (gebaseerd op microstappen voor de contourering van tumoren) tot een gemiddelde afname met 16% van de resulterende GTV-contouren. Daarom moeten ontwerpers toch terughoudend zijn met het ontwerpen van software die nauw aansluit op de werkflow, aangezien dit gevolgen kan hebben voor het sensemaking-proces en daarmee voor het klinische resultaat. Op basis van dit onderzoek is geconcludeerd dat de volgende stappen van het onderzoek de workflow van GTV-contourering volgen zoals dit gebeurt in de huidige klinische praktijk.

Om meer inzicht te kunnen krijgen in het sensemaking-proces en die inzichten te vertalen in ontwerpvereisten, is er een benadering in twee stappen voorgesteld. Deze benadering wordt gepresenteerd in paragraaf 5.2. De benadering bestaat uit: (1) het modelleren van sensemaking in context en (2) diepgaande analyse van software-interactie (patronen) met betrekking tot dat model. Deze benadering is gehanteerd bij het bestuderen van het maken van contouren van tumoren door acht artsen met behulp van het softwareprototype. De interacties met de softwareprototypes zijn vastgelegd in logbestanden. Deze interactielogs zijn vervolgens gevisualiseerd als tijdlijnen (paragraaf 5.3), op basis waarvan conclusies zijn getrokken over sensemaking.

De contoureertaak bleek drie hoofdfases te hebben: vertrouwd raken, actie en evaluatie. Tijdens elke fase zijn interactiepatronen geïdentificeerd. Zo bleek ‘lange navigatie’ (een groot aantal 2D-doorsnedes bekijken) hoofdzakelijk plaats te vinden tijdens het vertrouwd raken en de evaluatie. Vervolgens zijn de geïdentificeerde interactiepatronen gekoppeld aan sensemaking-activiteiten. Zo waren bijvoorbeeld tijdens het vertrouwd raken de voornaamste sensemaking-activiteiten het opbouwen van een initieel kader en verkennend naar informatie zoeken, terwijl de arts tijdens de evaluatiefase voornamelijk bezig was met gericht informatie zoeken. Op basis van die bevindingen zijn vijf punten van verbetering geïdentificeerd om het sensemaking-proces beter te ondersteunen: (1) mogelijk maken om effectief een initieel kader te ontwikkelen; (2) intuïtieve navigatie binnen en tussen datasets ondersteunen; (3) de detectie van relevante gebieden ondersteunen; (4) mogelijk maken om andere methoden voor het evalueren van contouren toe te voegen en (5) de algemene efficiëntie verbeteren door de benodigde tijd en fysieke inspanning te verkleinen. De conclusie is dat de voorgestelde tweestaps benadering gunstig is voor het verkrijgen van gedetailleerd inzicht in het sensemaking-proces en het bepalen van ontwerpvereisten die sensemaking ondersteunen.

Efficiënter gebruik van de tijd bleek inderdaad een belangrijke uitdaging te zijn bij het ontwerpen van software om sensemaking te ondersteunen. In hoofdstuk 4 is de mogelijkheid om automatisering in te bouwen al vanuit een meer technische invalshoek verkend. Maar met betrekking tot het gebruik van automatisering in een softwareoplossing om sensemaking te ondersteunen was niet duidelijk wat voor invloed automatisering kan hebben op het redeneerproces en/of de uiteindelijke contourering. Daarom is verder onderzoek verricht om de invloed van een veelgebruikt middel voor automatische contourering, namelijk het interpoleren van kleur tussen doorsnedes, te analyseren. Het bleek dat artsen die dit hulpmiddel gebruiken de taak efficiënter konden uitvoeren, omdat zij de initiële kleur niet meer hoefden in te tekenen. Daarnaast werd opgemerkt dat de resultaten van interpolatie tussen artsen consistenter waren. Tegelijkertijd had die automatisering ook invloed op het contoureringsproces en er werd minder tijd besteed aan het inspecteren van de data en resultaten.

In dit proefschrift wordt onderzoek gepresenteerd dat is uitgevoerd naar ontwerpen voor sensemaking. Dit levert twee inzichten op: kennis over het sensemaking-proces van artsen tijdens het maken van contouren en een nieuwe benadering voor het ontwerpen van software die het maken van contouren ondersteunt. Dit onderzoek moet worden beschouwd als de eerste stap in de richting van software die de sensemaking door artsen tijdens het maken van contouren van tumoren ondersteunt. Daarnaast is de voorgestelde benadering, waarbij het perspectief van sensemaking wordt meegenomen, ook bedoeld voor gelijksoortige ontwerpuitdagingen.

READING GUIDE

This thesis encompasses the disciplines of cognitive ergonomics and radiotherapy and is intended for both software designers and physicians. The following sections provide a brief overview of the relevant concepts and terminology in these two fields.

Cognitive ergonomics

Cognitive ergonomics and interaction design

Cognitive ergonomics is concerned with mental processes that affect interactions among humans and other elements of a system (International Ergonomics Association (IEA) 2017). One of the key domains in which cognitive ergonomics is applied is Human-Computer Interaction (HCI). HCI is the study of the way in which computer technology influences human work and activities (Blanton et al. 2009). Interaction design is the practice of designing interactive (digital) products (Cooper et al. 2007), covering both the function (i.e. how it works) and the form (i.e. how it looks). The terms HCI and interaction design are often used synonymously depending the intended area of emphasis.

Usability

The ISO/IEC 62366 Medical Devices - Application of Usability Engineering to Medical Devices (Part 1) standard defines usability as the “characteristic of the user interface that facilitates use and thereby establishes effectiveness, efficiency and user satisfaction in the intended use environment” (International Organization for Standardization 2015).

Sensemaking

This work focuses on the cognitive process of sensemaking: the process of creating understanding to inform action (Zhang and Soergel 2014). Since this task involves a range of uncertainties, sensemaking is the key underlying cognitive process. Throughout this process, the sensemaker continuously cycles between the (cognitive) frames (i.e. representations, structures, mental models and knowledge) and data, which results in data being re-framed and updated.

Medical context

This thesis is rooted in the context of radiotherapy.

Medical imaging

Medical imaging is the technique and process that seeks to reveal the internal structures of the human body. The outcome from medical imaging is a set of medical images representing the 3D human body. Physicians examine these images on 2D screens primarily in three different planes: axial, coronal, and sagittal as illustrated Figure I - 1. A 2D image presented on any give plane is typically referred to as a slice.

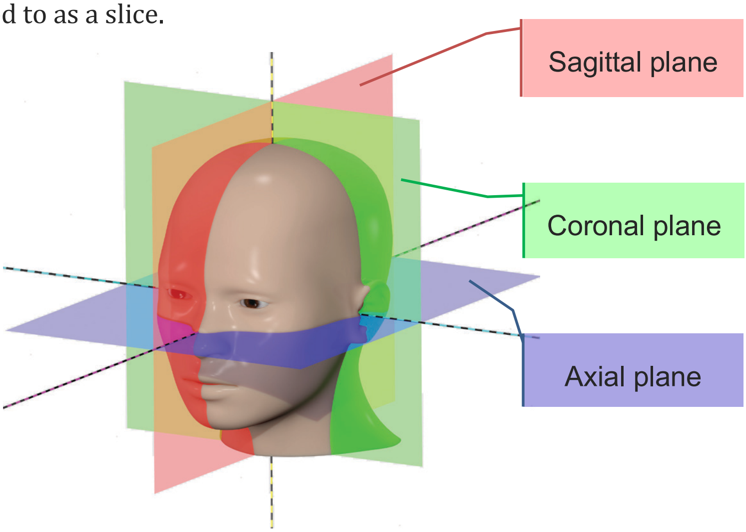


Figure I - 1 The three primary anatomic planes. Adapted from (Slashme 2014), licensed under CC BY-SA 4.0, via Wikimedia Commons.

Basic medical imaging modalities

There are different types of medical imaging techniques. Contrast enhancement in an image means that the patient was injected with a special chemical that made certain tissues more visible in the medical images (i.e. enhancing the quality of the images).

- Computed Tomography (CT) utilizes x-rays to create 3D images. CT may be used with or without contrast enhancements.
- Magnet Resonance Imaging (MRI) utilizes the principle that different tissue cells react differently to magnetizations making it possible to distinguish them from one another. There are different types of MRI images, such as:

  - MRI T1-weighted, which captures the longitudinal relaxation time and may be generated with or without contrast enhancement;
  - MRI T2-weighted, which captures the transverse relaxation time;
  - MRI FLAIR, which is similar to MRI T2-weighted but the signal of cerebrospinal fluid (CSF), which is fluid surrounding the brain and spine is suppressed.
- Positron Emission Tomography (PET) utilizes different types of radio-labeled molecules that accumulate in specific regions in the body making those regions 'visible'. For example, radiolabeled 2-fluoro-2-deoxy-D-glucose (FDG) cannot be metabolized by tumor cells and therefore accumulates within them, which enables those regions to be detected.

There are more modalities of medical imaging (e.g. ultrasound), but they are not addressed in this thesis.

Tumor

A tumor (i.e. malignant cancer) is a growth of abnormal tissue in the human body. This thesis is primarily concerned with a very aggressive type of primary brain tumor known as Glioblastoma Multiforme (GBM). GBMs usually occur in the cerebral hemispheres of the brain (see Figure I - 2). In current clinical practice, radiotherapy forms a standard part of GBM treatment (Stupp et al. 2009).

Radiotherapy

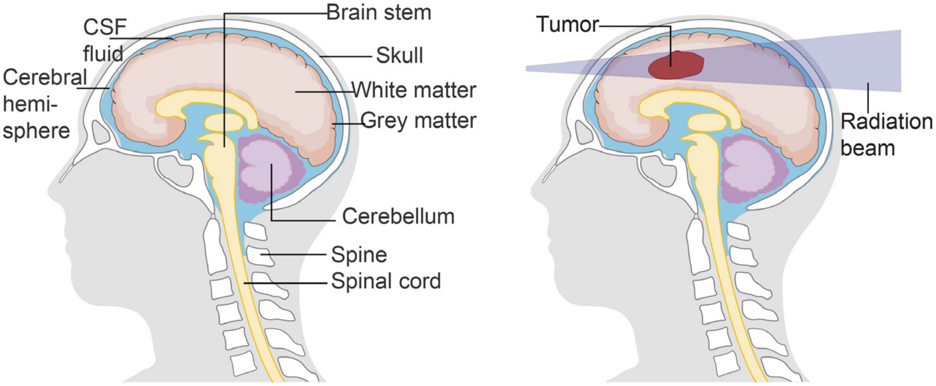


Figure I - 2 Basic structure of the brain (left). Illustration of photon beam radiotherapy (right). Own work based on (Cancer Research UK uploader 2016)(Licensed CC BY-SA 4.0)

Radiotherapy is a type of treatment which utilizes radiation to break down the structure of tumor cells and consequently kills them. Radiotherapy aims to deliver high doses of radiation to the tumor (Figure I - 2), while attempting to spare healthy tissue (i.e. organs at risk (OARs)) (Burnet 2004).

Medical imaging of brain tumors for radiotherapy

In order to plan radiotherapy treatment, physicians generate medical images of the patient. The types of images acquired depend on the specifics of the tumor and the practices of that particular medical institute (Batumalai et al. 2016). There are two main types of medical images of brain tumors that are captured for radiotherapy purposes (Drevelegas and Papanikolaou 2011): CT and MRI.

- CT provides information relating to tissue density and, in current clinical practice, is required in order to calculate the radiation dose (Pereira et al. 2014). It also provides useful information regarding bone.
- MRI T1-weighted: Tissues with high fat content (e.g. white matter) appear bright and compartments filled with water (e.g. CSF) appear dark (Mader 2015). MRI T1 contrast enhanced images show a higher level of contrast between different types of tissue, making it easier to detect abnormalities.
- MRI T2-weighted: Compartments filled with water (e.g. CSF compartments) appear bright and tissues with high fat content (e.g. white matter) appear dark. Most (but not all) lesions are associated with an increase in water content (Mader 2015).
- MRI FLAIR: This technique produces images similar to MRI T2 except that signal of the CSF is suppressed, rendering it dark on images instead of bright.

The use of PET images for brain tumors is not yet widely used in standard practice (Chen 2007).

Target volumes

The region to which the radiation dose needs to be delivered is defined based on volumes identified in medical images. A volume is created as a stack of 2D contours (see Figure I - 3 left). There are three primary volumes for GBM tumors (see Figure I - 3 right):

- Gross Tumor Volume (GTV): represents the ‘visible’ tumor;

- Clinical Target Volume (CTV): an expansion of GTV that includes the ‘invisible’ spread of the tumor cells identified from clinical research;
- Planning Target Volume (PTV): an expanded volume of CTV that takes account of possible movements during the delivery of the treatment.

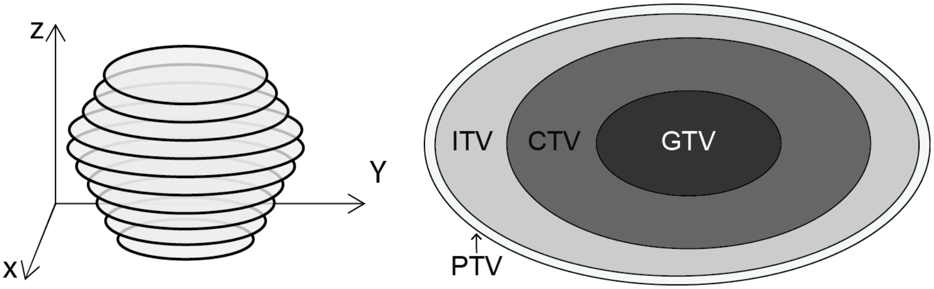


Figure I - 3 An illustration of a stack of 2D contours representing a 3D volume (left), and the relationships between the three primary volumes in 2D view (right)

Near-synonyms

During the course of this research project, we identified that some terms are used synonymously depending on the context. While we acknowledge the differences between these terms, depending on the context they may not be relevant.

- Contouring / delineating / delineation of / segmenting / segmentation of = the process of drawing the contours;
- Segmentation method / segmentation algorithm = computational contouring;
- Image registration / image co-registration / image fusion = the process of aligning two medical image datasets to the same coordinate space.

List of acronyms

Table I - 1. List of acronyms

Acronym	**Description**
2D	Two-dimensional
3D	Three-dimensional
BHD	Bidirectional Hausdorff Distance
BMHD	Bidirectional Mean Hausdorff Distance
CE	Contrast Enhancement, Contrast Enhanced
CSF	Cerebrospinal fluid
CT	Computed Tomography

CTV	Clinical Target Volume, the area surrounding a GTV suspected of being tumorous. Typically obtained by applying a margin to the GTV
DJC	Dice Jaccard Coefficient
GBM	Glioblastoma Multiforme
GTV	Gross Tumor Volume, the ‘visible’ tumor
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HD	Hausdorff Distance
ISO	International Organization for Standardization
MRI	Magnet Resonance Imaging
OAR	Organ at Risk. Organ (i.e. healthy tissue) in proximity to tumor
PET	Positron Emission Tomography
PTV	Planning Target Volume, the area for which the radiation dose is planned. Typically obtained by applying a margin to CTV
ROI	Region of Interest (e.g. GTV, CTV, PTV, OAR)
RT	Radiotherapy
SD	Standard Deviation
SUMMER	Software for the Use of Multi-Modality images in External Radiotherapy
WF	Workflow

Chapter 1

Introduction

This chapter presents an overview of the topic, the problem statement, the research focus and the research approach of this Ph.D. project as well as an outline of this thesis.

1.1 Radiotherapy

Cancer incidence is expected to increase due to the aging population (Yancik and Ries 2004). Among the various types of treatments, radiotherapy is one of the most effective methods for treating cancer (Njeh 2008). Radiotherapy uses high energy radiation to shrink tumors by destroying cancer cells (National Health Service (UK) 2015). It works by damaging the DNA of the tumor cells, which are then unable to reproduce. Delaney et al. (2005) estimated that 52 per cent of cancer patients can potentially benefit from this treatment. In the Netherlands, approximately 48 per cent of cancer patients receive radiotherapy treatment (Slotman and Leer 2003; Grau et al. 2014). Radiotherapy treatment may also improve overall survival rates. For example, in a recent study, Corradini et al. (2015) indicated that in the context of breast cancer management, ten-year overall survival rates were 82 per cent following postoperative radiotherapy ($p < 0.001$) as opposed to 55 per cent with surgery alone.

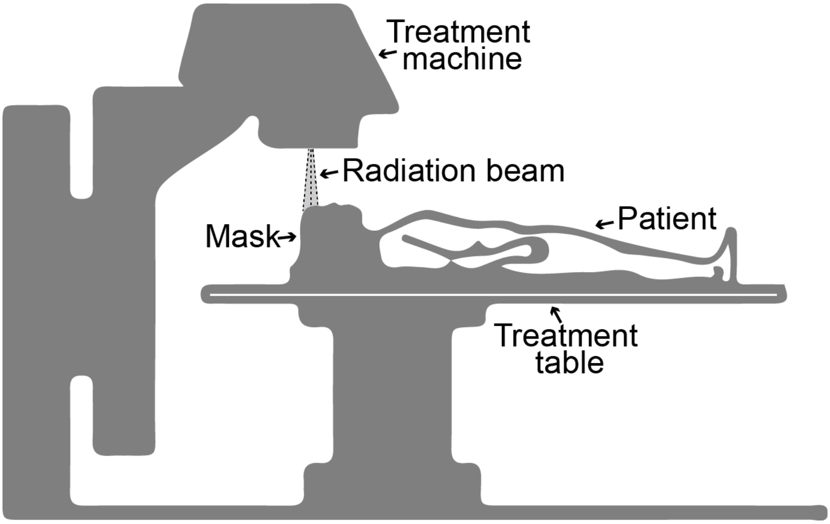


Figure 1-1 Illustration of the external radiotherapy treatment set-up for brain tumor treatment. The patient lies on the treatment table, the head is fixed with a mask and the treatment machine delivers the radiation.

Radiotherapy can be delivered either internally, where the source of radiation is placed inside or near the tumor, or externally, where the radiation is delivered as a beam or beams of high-energy X-rays (Kirthi Koushik et al. 2013). During external radiotherapy treatment, the patient must be positioned precisely on the treatment table, and the treatment machine (a linear particle accelerator) directs the radiation beams to the defined location in accordance with the dosage plan (see Figure 1-1.).

The radiotherapy workflow consists of four main phases (Aselmaa et al. 2013c): diagnosis, radiotherapy treatment planning, treatment delivery, and post-treatment follow-up (Figure 1-2). Each of these phases has its own challenges. For instance, timely diagnosis improves patients' chances of survival. However, in many cases patients delay seeking help (Smith et al. 2005). Like diagnosis, the main challenge for post-treatment follow-up is early detection of possible relapses. The challenges facing treatment planning range from acquiring good-quality medical images to creating an optimal radiation dosage plan. The challenge for treatment delivery is to ensure the correct positioning of the patient and the precise delivery of the radiation (e.g. Figure 1-1). This thesis focuses primarily on the radiotherapy treatment planning phase.

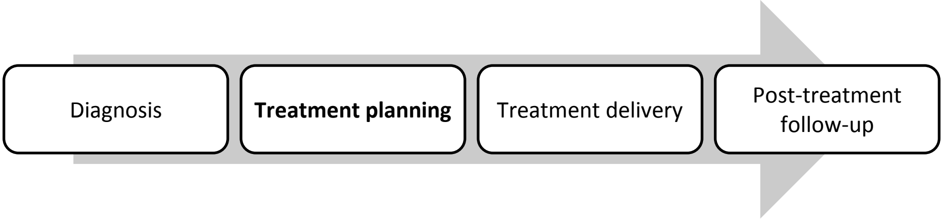


Figure 1-2 General workflow for radiotherapy.

The treatment planning phase commences with acquiring medical images of multiple modalities (e.g. CT, MRI and PET) as deemed necessary by the physician, depending mainly on the type of tumor (Batumalai et al. 2016). Each imaging modality provides unique clinical information relevant for planning treatment. Images from different modalities are then co-registered in the same coordinate space to facilitate the extrapolation of information in the same location. The relevant regions are subsequently contoured (i.e. the tumor and nearby organs). Once all the relevant contours have been created, the radiation dosage is planned and the overall treatment plan validated (e.g. Winkel et al., (2016)). Further details of this process are explained in Chapter 2.

Contouring is one of the critical steps in treatment planning. During the contouring task, the physician contours the treatment volumes as well as the healthy surrounding tissue (Vieira et al. 2016). This involves drawing the visible borders of the tumor or organ on a number of slices (Dowsett et al. 1992) resulting in a set of 2D contours representing the 3D volume. Different types of volumes are required for treatment planning (Purdy 2004). For example, the Gross Tumor Volume (GTV) represents the tumor that is visible on medical image datasets and also palpable during physical examination (Burnet 2004). Other volumes are then identified based on the GTV by incorporating medical knowledge regarding the expected tumor spread (i.e. Clinical Target Volume (CTV), which incorporates the non-visible areas of the tumor), and uncertainties surrounding the treatment delivery (i.e. Planning Target Volume (PTV), which takes account of possible movements of the patient).

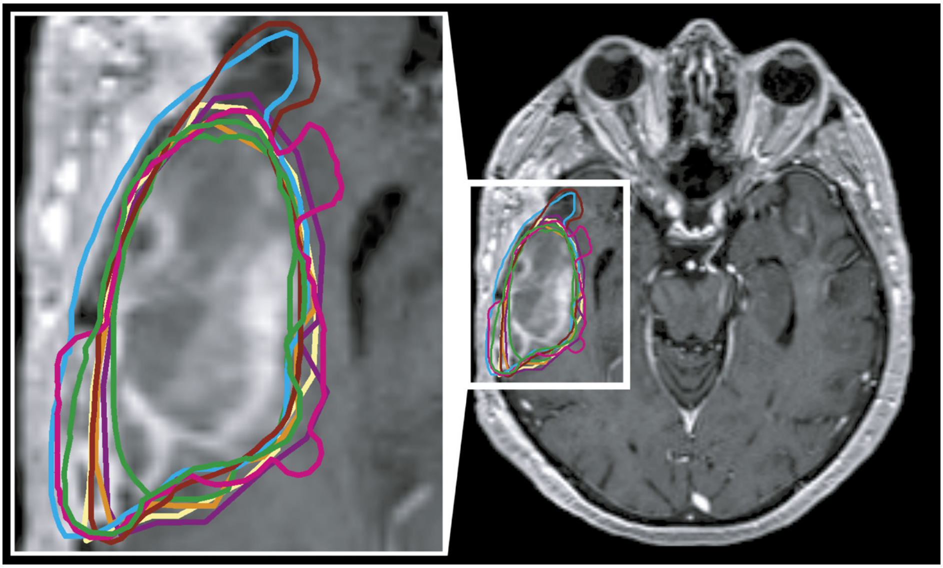


Figure 1-3 Example of variation in GTV contours (published in (Aselmaa et al. 2017)

Technological advances in the past decades have made it possible to deliver radiation to areas of highly-complex shape (Nutting et al. 2000). It is crucial to identify all the relevant volumes accurately and efficiently for optimal radiotherapy treatment. However, tumor contouring has been considered the weakest link in radiotherapy planning (van Herk 2004; Njeh 2008). Tumor boundaries on medical images are often unclear, which makes it challenging for physicians to distinguish between tumorous tissue and normal tissue. Moreover, different imaging modalities provide different types of information which may be conflicting. Physicians also need to consider other variables, such as treatment details and tumor characteristics (Aselmaa et al. 2014). Therefore, throughout the contouring task, physicians need to obtain and synthesize different types of data along with their knowledge and experience in order to reach a decision. Consequently, physicians' interpretations of the data and the contours they identify may vary, as illustrated in Figure 1-3.

1.2 Research context

This research project was completed in the Faculty of Industrial Design Engineering at Delft University of Technology (TU Delft), and was initiated as a part of a European research project. This setting resulted in the research focusing on radiotherapy and interface and interaction design.

1.2.1 SUMMER project

The author was a member of the Software for the Use of Multi-Modality images in External Radiotherapy (SUMMER) project research team, which was part of the Marie Curie Research Training Network (PITN-GA-2011-290148) and funded by the European Commission's Seventh Framework Programme. This SUMMER software was intended to contribute to clinical efforts to better target tumorous tissue while increasing safety for normal tissue (2017a).

The SUMMER project was completed by a multi-disciplinary team formed at the beginning of the project in 2012. The team consisted of seven members: three hospitals (Universitätsklinikum Freiburg (Germany), Fondazione Santa Lucia (Italy) and Institut Claudius Regaud (France)); two companies (Aquilab (France) and VRVis (Austria)); and two universities (Medical University of Vienna (Austria) and TU Delft (The Netherlands)). Each member hosted one or two researchers in addition to the scientist in charge. The members also had expertise in a specific area and were assigned an associated task accordingly. TU Delft was responsible for providing the consortium with information regarding users and user-system interaction, conducting investigations with end-users, and informing the design of the software prototype. I myself worked closely with the Universitätsklinikum Freiburg and Institut Claudius Regaud hospitals.

1.2.2 Interaction design

Interaction design is the practice of designing interactive (digital) products, environments, systems and services (Cooper et al. 2007). These interactive products may be digital only (i.e. software) or combined with a physical product (e.g. a smart television). This research concentrated on the interaction design of a software system and was guided by the principles of human-centered design.

Section 210 of ISO standard 9421 lists requirements and recommendations for human-centered design principles and activities throughout the life cycle of computer-based interactive systems (ISO 9421-210: 2010). One of these principles is that "*the design is based upon an explicit understanding of users, tasks, and environments*" (i.e. designers must understand the context in which the product is used). In this work, we refer to this step as the 'early phase of design', where software projects typically commence. However, human-centered design is an iterative process, therefore activities to increase our understanding are expected to occur at every iteration.

One important aspect of the context of use is understanding the users' cognitions. For example, the US National Research Council has published a set of principles to guide successful utilization of IT in healthcare to support a 21st century vision of healthcare (Stead et al. 2009). One of those principles was supporting the cognitive functions of all caregivers (Stead et al. 2009). Designers should therefore increasingly focus on understanding and designing for cognitive processes.

1.2.3 Cognitive informatics

Cognitive informatics is an emerging interdisciplinary field that draws on psychology, computer science, cognitive science and other areas to understand human activities such as reasoning, decision-making and problem-solving (Patel and Kannampallil 2015). Problem-solving and decision-making are two key paradigms for psychological research on clinical reasoning (Elstein and Schwartz 2002), in which clinical problem-solving involves selecting a hypothesis: *"solutions to difficult diagnostic problems were found by generating a limited number of hypotheses early in the diagnostic process and using them to guide the subsequent collection of data"* (Elstein 1978). Clinical decision-making is also viewed as a form of opinion revision: *"reaching a diagnosis means updating opinion with imperfect information (the clinical evidence)"* (Elstein and Schwartz 2002).

The need to support physicians' cognitive processes has attracted attention in recent decades (Patel and Kannampallil 2015). One main application area is designing systems that support clinical decision-making by *"delivering one or more specific pieces of clinical knowledge or data to an individual at a specific time and place"* (Osheroff et al. 2004). Such systems are suitable for well-defined clinical problems, for instance, by providing reminders for physicians regarding certain tasks (Kawamoto et al. 2005).

‘Decision support system’ is an umbrella term for many different types of intervention systems. For tasks where the process or judgements on the ‘correctness’ of the outcome are somewhat fuzzy, sensemaking can be seen as the underlying cognitive process. ‘Sensemaking supportive software’ can therefore be defined as a sub-category of decision support system that supports sensemaking.

1.2.4 Sensemaking

The notion of sensemaking has emerged in recent years to describe a cognitive process that people engage in when managing uncertainties and fuzziness in a given task. Klein et al. (2006a) described sensemaking as a motivated, continuous effort to understand connections (e.g. relationships) between people, places and events etc. in order to act effectively. In workplaces where software is central to task completion, users can explore possible connections, investigate hypotheses, and ultimately gain insights through interactions with the software (Endert et al. 2012). Investigating physicians’ cognitive processes while completing tasks using software from a sensemaking perspective is a promising avenue for improving software design.

Research into sensemaking has been conducted since the 1980s in four main fields (Dervin and Naumer 2009): library and information science (Dervin 1998), organizational communication (Weick 1995b), human-computer interaction (HCI) (Russell et al. 1993), and cognitive systems engineering (Klein et al. 2006b). In the following sections, we present an overview of these sensemaking theories.

1.2.4.1 Dervin’s Sense-Making approach

> *“Sense-Making [..] is defined as behavior, both internal (i.e. cognitive) and external (i.e. procedural) which allows the individual to construct and design his/her movement through time-space.”* (Dervin 1983)

Dervin’s Sense-Making approach studies users to understand them and design systems to meet their needs. She developed it to focus on users’ “sense making” and “sense unmaking” in the fields of communication, and library and information science. The approach assumes that humans “live in a world of gaps: a reality that changes across time and space and is at least in part ‘gappy’ at a given time-space” (Dervin 1998). In Dervin’s view, the term "Sense-Making" is “a label for a coherent set of concepts and methods [...] to study how people construct sense of their worlds and, in particular, how they construct information needs and uses for information in the process of sense-making [behavior]" (Dervin 1983).

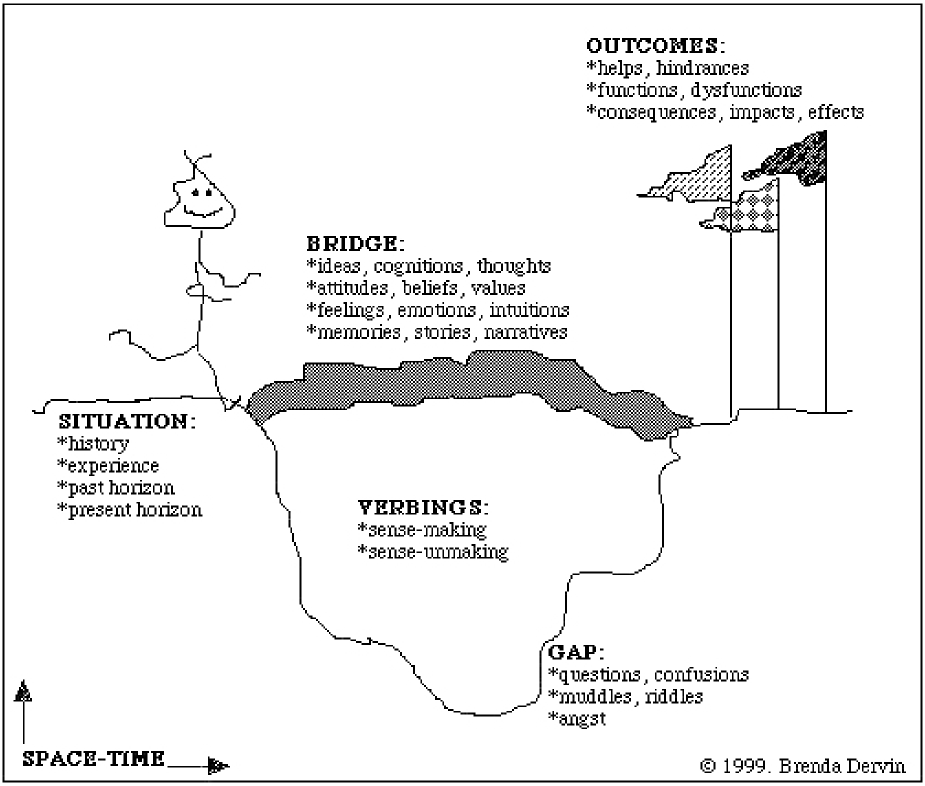


Figure 1-4 Sense-Making Metaphor illustrated by Dr B. Dervin (via Wikimedia Commons)

Figure 1-4 illustrates the Situations-Gaps-Uses model used in Dervin's Sense-Making approach studies (Dervin 1983). 'Situations' are defined as the time-space context at which sense is constructed. 'Gaps' are seen as needing bridging, translated in most studies into 'information needs' or the questions people have as they construct sense and move through time-space. 'Uses' are defined as the ways individuals use the newly created sense, translated in most studies as 'information helps' and 'hurts'.

1.2.4.2 Sensemaking within organizations

> *"Sensemaking is what it says it is, namely, making something sensible."*
> (Weick 1995b)

In his book Sensemaking in Organizations, Weick (1995b) defines sensemaking as “making of sense” according to seven characteristics:

1. Grounded in identity construction: Weick states that one sensemaker is a “parliament of selves” and consequently, depending on who the sensemaker is at the moment of sensemaking, the sensemaker’s definition of what is out there will also change.
2. Retrospective: An action can become an object of attention only after it has occurred.
3. Enactive of sensible environments: The environment cannot be separated from the sensemaking since people often produce part of the environment they face.
4. Social: Human thinking and social functioning are essential aspects of each another, therefore sensemaking is also a way of being social.
5. Ongoing: People are always working on something, therefore sensemaking is an ongoing process.
6. Focused on by extracting cues: Extracted cues are simple, familiar structures that are the seeds from which people develop a larger sense of what may be occurring.
7. Driven by plausibility rather than accuracy: Sufficiency and plausibility take precedence over accuracy – in other words “accuracy is nice, but not necessary” (Weick 1995b)

1.2.4.3 Learning loops in sensemaking

> *“Sensemaking is the process of searching for a representation and encoding data in that representation to answer task-specific questions.”*
> (Russell et al. 1993)

Russell et al. (1993) identified sensemaking through the core process ‘learning loops complex’ which consists of four main aspects, which are illustrated in Figure 1-5 where the term ‘representation’ is used as a synonym of ‘frame’ or ‘structure’.

1. Search for representations: This search is the generation loop. The sensemaker creates representations that capture some notable features of the data so that it is in line with the instantiated representation.
2. Instantiate representations: Instantiated representations are called ‘encodons’ and are created in the data coverage loop. The sensemaker

repeatedly identifies information of interest and encodes it in a representation that has emerged from the generation loop.

3. Shift representations: The representational shift loop is guided by the discovery of residue, i.e. ill-fitting or missing data and unused representations. These shifts during sensemaking are intended to reduce the cost of task operations. When there is relevant data that has no place in the representation, the representation can be expanded. When data does not fit the established categories, the original representation categories may need to be merged, split, or new categories added.
4. Consume encodons: The sensemaker then uses the encodons in a task-specific information processing step.

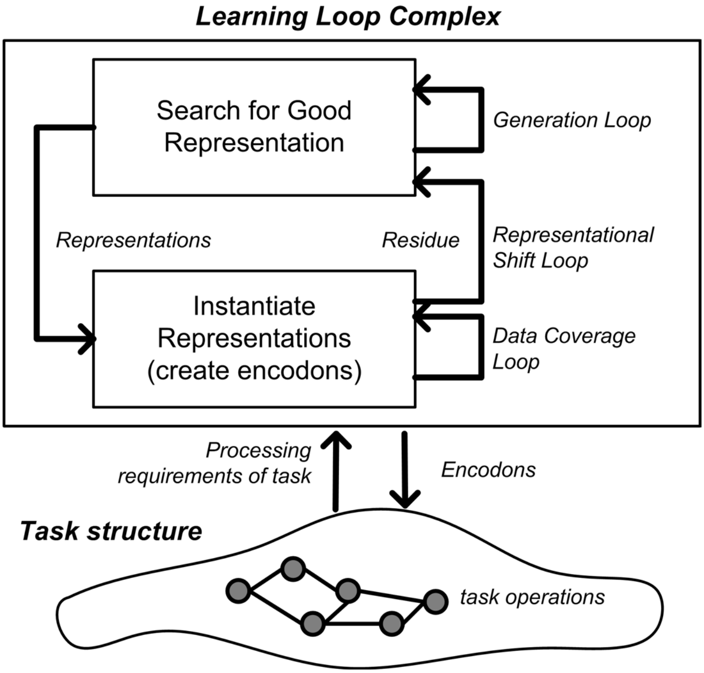


Figure 1-5 The learning loops of sensemaking as modeled by Russel et al. (1993), adapted from (Russell et al. 1993)

“If there were no surprises in creating encodons, sensemaking would be trivial; merely define the schemas and then instantiate them. Sensemaking seldom works this way.” (Russell et al. 1993). Therefore, sensemaking requires revising representations when the sensemaker encounters surprises while creating encodons, or as new task requirements come to light.

1.2.4.4 Data/Frame Theory of sensemaking

> *"Sensemaking is a motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively."* (Klein et al. 2006b)

Klein et al. (2006b) describe sensemaking as a symbiosis between data and frame (i.e. representation and structure) as illustrated in Figure 1-6. A frame functions as a hypothesis about the connections between data and the sensemaker's knowledge. During the sensemaking process, doubts may arise regarding the frame or data which can then either be explained away while preserving the frame or the frame can be 'elaborated' to incorporate the new data. These two aspects, preserving the frame and elaborating the frame, form part of the elaboration cycle of sensemaking (see Figure 1-6, left). Another cycle of sensemaking is to reframe (see Figure 1-6, right). Here, questioning the frame leads the sensemaker to reconsider the current frame and replace it with a new one. During re-framing, the sensemaking activity comprises finding a frame that plausibly links the events requiring explanation.

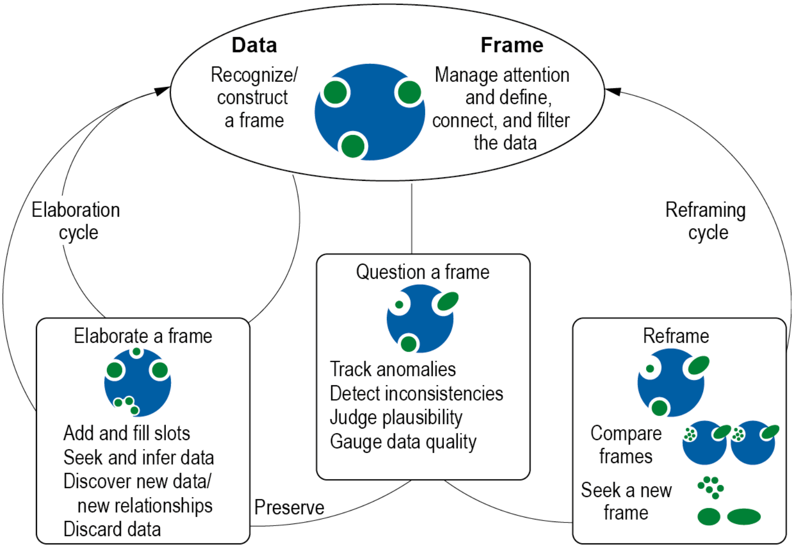


Figure 1-6 Data/Frame theory of sensemaking as presented by (Klein et al. 2006b)

1.2.4.5 Revised model by Zhang et al., 2014

> "Sensemaking is the information task of creating an understanding of a concept, knowledge area, situation, problem, or work task […] often to inform action." (Zhang and Soergel 2014)

During the period in which this research project was being completed, Zhang et al. (2014) published a review of sensemaking theories which discussed details, similarities and differences between various sensemaking perspectives. In addition to the theories described above, the authors also incorporated learning theories and models from cognitive psychology. They proposed a comprehensive model of sensemaking that combines concepts and techniques from different but closely linked fields (see Figure 1-7).

In this proposed model, the sensemaking process consists of several iterative loops of information seeking and sensemaking. The start point for the sensemaking process is the sensemaker's existing knowledge (or lack thereof) of the problem or the work task situation and the end point is an updated conceptual structure that is iteratively updated through accretion, tuning or restructuring. The information seeking activities are either exploratory for data or focused for the structure.

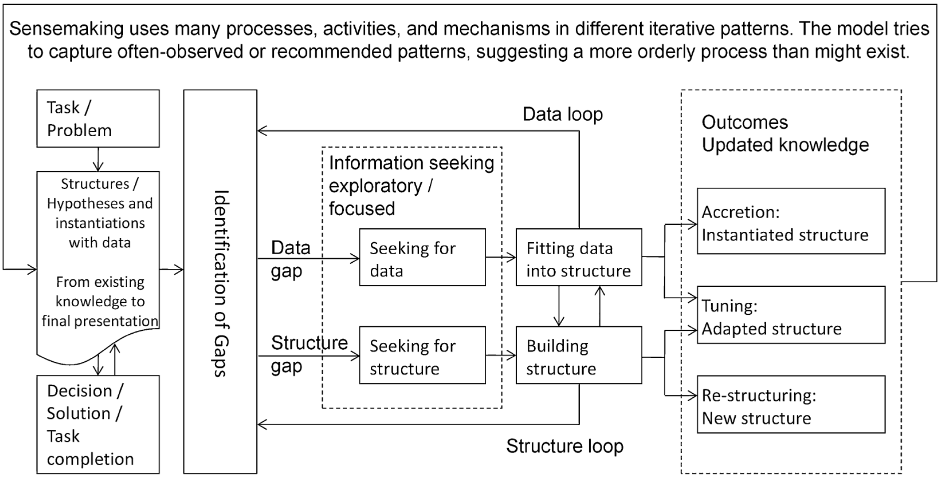


Figure 1-7 The model of the cognitive process and mechanisms of individual sensemaking as proposed by (Zhang and Soergel 2014)

1.2.4.6 Summary of sensemaking theories

Our review of the relevant literature indicated that the models of sensemaking process vary somewhat depending on the interpretation of the researchers and the field of application. However, sensemaking is generally viewed as an iterative process that starts with a defined goal, and uses data to iteratively build and update frames (i.e. knowledge, mental models) until the sensemaker reaches a satisfactory outcome. Furthermore, gaps (i.e. discrepancies between data and frames, or between frames) are typically understood to as the triggers of sensemaking activities. The driving force for sensemaking activities is to explain gaps, which result in the frames or data being updated. Therefore, taking a broad view, sensemaking connects the data and the associated frame through a series of sensemaking loops to build and update the frame according to a specific task goal (Aselmaa et al. 2017).

1.3 The gap

To date, research into the contouring task has primarily been concerned with increasing time efficiency (e.g. by incorporating automation) since the contouring process can be lengthy and tedious (Dowsett et al. 1992). The next area for research is on how to assist physicians to reach an optimal contour by supporting their cognitive processes. This requires a deep understanding of these processes and sensemaking may be considered the underlying process in which physicians are engaged during the cognitively demanding task of contouring.

It is anticipated that a software system that is in line with physicians' cognitive processes will enhance their performance. Nevertheless, from an interaction design perspective, the question of how to incorporate the theoretical notion of sensemaking into the design process remains to be answered. The challenge in this regard is the range of different theories and models concerning sensemaking.

1.4 Research focus and approach

The aim of this research project is:

> *To provide the means to understand physicians' sensemaking process during the early phase of design in order to design software that is well-fitted to the clinical workflow.*

The research presented in this thesis explores ways of designing to take account of sensemaking by conducting a series of studies to describe physicians' sensemaking process during the tumor contouring task, while identifying the design requirements for sensemaking support software. The scope of this research incorporates three topics: radiotherapy, interaction design, and cognition (see Figure 1-8). We address the following research questions (RQs):

- RQ-1: What is the workflow in radiotherapy?
- RQ-2: What are the cognitive processes involved in the contouring task? Which medical factors influence contouring?
- RQ-3: What are the challenges for incorporating automated contouring into software design?
- RQ-4: How can we incorporate sensemaking theory into the early phase of software design?
- RQ-5: What sensemaking process do physicians follow during tumor contouring?
- RQ-6: How does automated contouring influence physicians' cognitive processes?

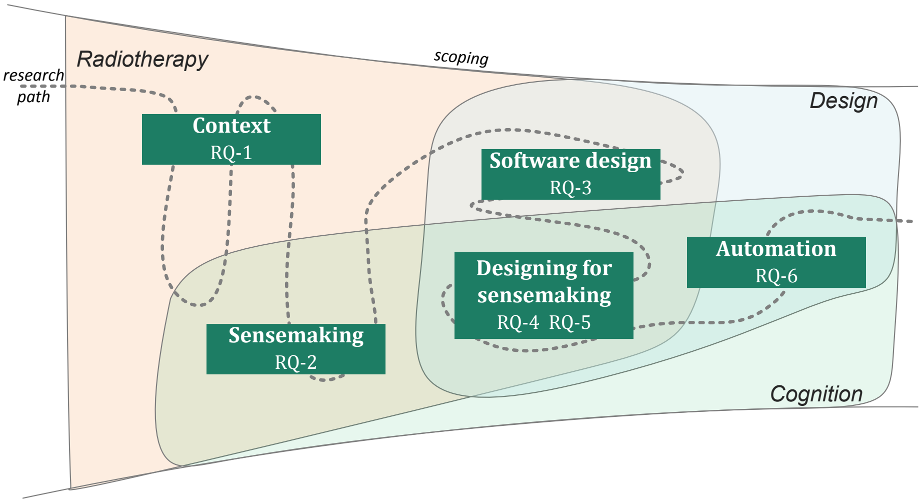


Figure 1-8 Our research approach in relation to our research questions (RQ).

It is worth mentioning that we used a range of different research and design methods throughout this research project, including human factors and ergonomics methods (e.g. workflow analysis, observations, interviews (Stanton et al. 2005)), co-design methods (e.g. workshops (Freudenthal et al. 2011), collaborative prototyping (Sanders and Stappers 2014)), rapid software prototyping (Joseph 2004; Sass and Oxman 2006), and action research (Avison et al. 1999).

1.5 Thesis outline

The structure of this thesis follows that of the research approach described in previous section. It is set out in seven chapters (see Figure 1-9).

Chapter 1 (this chapter) introduces the topic and the theory of sensemaking.

Chapter 2 presents the outcomes of our radiotherapy workflow analysis and describes the main tasks and the stakeholders involved. We then elaborate on the treatment planning phase of the workflow in more detail. Finally, we highlight key areas for improvement regarding software design, as well as the implications of recent technological and clinical developments.

Chapter 3 outlines the cognitive aspects of the contouring task and identifies a number of medical factors that influence physicians' reasoning.

Chapter 4 addresses the challenges of incorporating automated contouring into radiotherapy software and describes the software prototype used in our studies.

Chapter 5 reports on two studies on software design for radiotherapy from a sensemaking perspective. In Section 5.1, we elaborate on our explorations of tumor contouring workflows in relation to software design. In Section 5.2, we propose a two-step approach for integrating sensemaking theory into the software design process and describe the results of applying this approach to the tumor contouring task. Section 5.3 subsequently presents a user interaction data analysis tool that we used to analyze the study results and to comprehend physicians' sensemaking.

Chapter 6 sets out the outcomes of our study on the influence of automatic contouring on physicians' cognitive processes.

Chapter 7 discusses the outcomes of the research project as a whole. We summarize the key requirements for contouring software that supports sensemaking. We also discuss the limitations of this project and finally highlight future research opportunities.

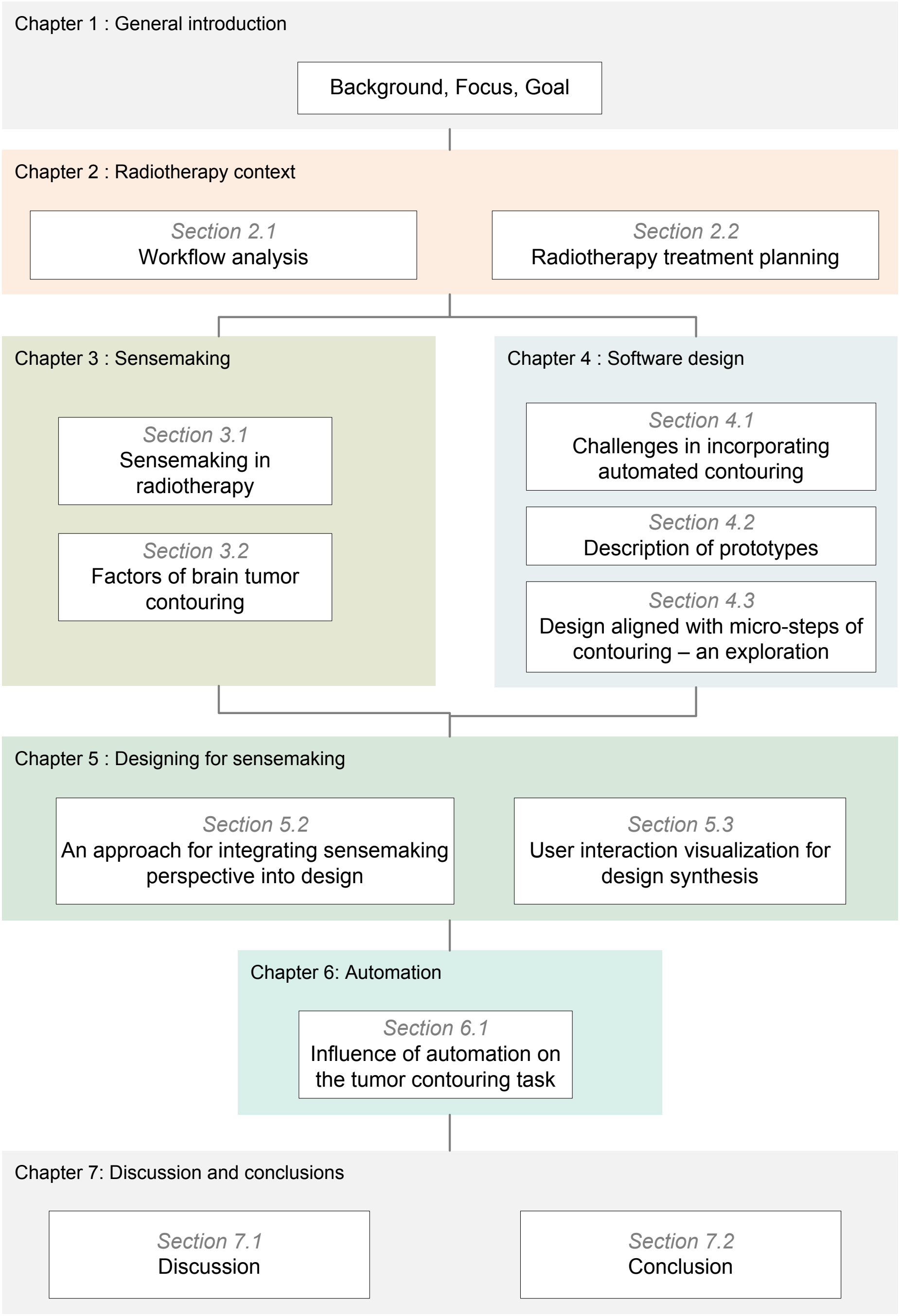


Figure 1-9 Structure of this thesis

Chapter 2

Context of radiotherapy

This chapter is based on:

Aselmaa A., Goossens R.H.M., Laprie A., Ken S., Fechter T., Ramkumar A., & Freudenthal A. (2013). Workflow Analysis Report. http://summer-project.eu/work/deliverables/.

Aselmaa A., Goossens R.H.M., Laprie A, Ramkumar A, Ken S, Freudenthal A. External radiotherapy treatment planning–situation today and perspectives for tomorrow. In: MASSOPTIER L, VIARD R, editors. Innovative imaging to Improve Radiotherapy Treatment 1st ed. Lulu Enterprises; 2014. p. 91–8.

Analysis of the context is the first step in the design process. This chapter presents the results of analyzing the radiotherapy workflow. Section 2.1 gives an overview of the complete radiotherapy workflow. Further analysis of the treatment planning phase is given in Section 2.2. Graphical representation of the workflow is included at the end of the chapter. Key areas to improve in the software design, as well as the implications of the technological and clinical developments to the design, are highlighted.

2.1 Radiotherapy workflow

This work presents the workflow of radiotherapy. It was one of the deliverables of the SUMMER project (Aselmaa et al. 2013b). The aim of the workflow analysis was to identify key processes of the clinical practices and to create a general understanding of the topic. The presented workflow analysis was conducted primarily based on observational studies in two hospitals. Based on the analysis, a workflow diagram representing the tasks and involved people was created and is presented at the end of this chapter. The key areas for improvements from software design perspective were to improve the efficiency as well as to support the physicians in interpreting and understanding information.

2.1.1 Introduction

Understanding of the radiotherapy workflow is needed in order to design new software (and new User Interfaces (UI)) for human-computer interactions that fit into the clinical context. Understanding the workflow is especially important since there are also differences between institutions and/or countries in EU. Furthermore, understanding the workflow gives the multi-disciplinary team members in the consortium a basis to be able to work together regarding (1) what is happening in the medical procedure, and (2) what types of vocabularies are used in the communications. Based on such an understanding, the collaboration will be conducted in a more effectively and efficient manner.

In this section, an overview will be given of what is a workflow and roles of different people in the radiotherapy procedure. The entire radiotherapy workflow will be presented in flowcharts. Regarding some particular steps in the focus of the SUMMER project, e.g., contouring, a more detailed analysis will be presented in Chapter 4, 5 and 6.

In the design process, the workflow needs to be inspected iteratively for the goal and the task analysis, even to the level of cognitive micro steps (Cuijpers et al. 2012). The purpose is to change the workflow towards a new workflow, which incorporates a new design. Whether the changes in the workflow will be big or small, are yet to be seen.

Because of the development of new technologies and medical practice, the workflow cannot be seen as a static thing. More thorough analysis of new approaches of radiotherapy (e.g., adaptive radiotherapy, dose painting and proton therapy) and limitations of the current workflow are described by Aselmaa et al. (2013c). Design should anticipate or facilitate these expected trends as well.

Radiotherapy (radiation therapy) can be external (the radiation source is external to the patient's body) and internal. Radiotherapy uses high-energy radiation to kill cancer cells. It is often an effective way to kill cancer cells that cannot be removed during surgery (2010). Different from external radiotherapy, in internal radiotherapy, commonly named brachytherapy, the radioactive material is placed into the body in the proximity of the tumorous cells.

Table 2-1 Vocabulary

Term	**Definition within this thesis**
Workflow	Collection of linked tasks, resources and information elements which are involved in specific process to achieve a specific goal
Treatment	Refers to external radiotherapy treatment unless specified
Image co-registration	The step preceding image fusion where the different sets of data are transformed into one coordinate system. The result of co-registration aims to gather information from several image modalities and put together. Registered images may bring relevant/new information when fused together (i.e., overlaid).
Image fusion	The action where two (or more) different images are "merged" into one image for the presentation on the computer display. In daily informal communication among clinicians, often the term "image fusion" is used as a synonym to "image co-registration" or "image registration"
Dosimetry	The process of planning the dose distribution for radiotherapy treatment.
RT	Abbreviation for radiotherapy, within this thesis, refers to external radiotherapy.
Contouring	The process of identifying regions of interest (tumor or organs) by drawing a line on the border of the region of interest. Also, referred to as 'delineation' or 'segmentation'. 'Segmentation' typically refers to algorithm-based contouring with no or limited user involvement.

2.1.1.1 Workflow definition

Workflow can be defined as "the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules" (Hollingsworth and Hampshire 1994) or more loosely "the specific collection of tasks, resources and information elements involved in […] a circumstance comprise a workflow" (Basu and Blanning 2000). It can be even viewed that each alternative within a workflow creates a new workflow, or that each sub-process can be a stand-alone workflow. Within this thesis, the workflow is defined as "*collection of linked tasks, resources and information elements which are involved in specific process to achieve a specific goal*".

Figure 2-1 The observed workflow management approach at one of the hospitals. A folder represented the files of a patient; Arrows indicating the task flow.

Workflow management may be done with digital aids or physical aids. In one of the hospitals, at the time of the observations the treatment planning workflow was managed based on a "shelf system" (Figure 2-1). The patient was represented as a folder that contained the relevant information about the diagnosis and treatment. Each "box" on a shelf represented a task. The position of the patient folder indicated the status within the workflow, and arrows in the figure indicated the relations among different steps in the workflow.

2.1.1.2 Different workflows in radiation oncology

In a radiation oncology department, there are multiple workflows happening in parallel. They influence each other one way or another – failure or delay in one workflow (i.e., quality assurance (QA) of a treatment machine has not been finished on time) influences other (i.e., patient treatment has to be done on a different machine or rescheduled). The focus of this work is the general radiotherapy workflow of one patient.

Table 2-2 Different workflows in radiation oncology

Workflow	**Examples**
Administrative workflows	Scheduling patients; Ordering supplies
Machine QA workflows	Installing new machine; Daily QA; Weekly QA
Patient treatment workflows in external RT	Curative treatment or Palliative treatment
Clinical trial workflow	Trial specific

2.1.1.3 *Participants*

The process of creating and executing the external radiotherapy treatment plan spread over a long period of time and involves many participants. It is important to mention that there are international differences between the names of the professions and the tasks they are responsible for.

Table 2-3 The overview of (main) participants in external radiotherapy

Participant	**Definition**
Patient	The person with cancer
Patient's family	The supporting people who are accompanying the patient
Physician	A person with medical degree
Radiologist	A physician who is specialized in the interpretation and reading medical imaging
Technician/ Radiotherapy Technologist	A person who is skilled in using medical technology. In radiotherapy, the technicians are sometimes called as Technologist. They receive specific training for the different treatment machines/ accelerators
Radiation oncologist	A physician who is specialized in radiotherapy

Medical resident	A graduated medical student, who is in training in the clinical setting
Medical oncologist	A physician who is specialized on medication-based cancer treatment
Nuclear medicine physician	A physician, who diagnoses and treats different types of diseases using radioactive materials and techniques, also involved in PET image acquisition.
Surgeon	A physician who is qualified to perform surgery. The surgeons specialized in oncology are called as Surgical Oncologist
Medical physicist	A person who has finished a special training in medical physics
Dosimetrist	A person who is specialized in RT planning (from patient file management to images co-registration and dose plan computation)

2.1.1.4 Possible data within RT

- Imaging data
    - Diagnosis IMG*
    - Pre-operative IMG*
    - Immediate post-operative IMG*
    - RT planning IMG*
    - Follow-up IMG*
- Patient data
- Radiologist reports
- Surgeon's report
- Anatomopathological report
- Patient's history report
- Treatment protocols
- Clinical trial protocol
- Other clinical trial documents
- Delineation rules/guidelines
- ...

IMG* = CT, MRI*, PET or PET-CT

MRI* = Includes: MRI T1-weighted pre-contrast (before the injection of contrast agent), MRI T1-weighted post-contrast (after the injection of contrast agent), MRI T2-weighted, MRI FLAIR, MRI Diffusion, MRI Perfusion, MRI Spectroscopy (mono-voxel, multi-voxel), fMRI.

2.1.1.4.1 Data carrier means

There are many potential data carriers used in different situations.

- Paper - Patient folder; Other forms; Fax (clinical trial, reports from another hospital(s)).
- Digitized documents for the electronic patient folder - scans of medical reports from different departments (surgery, anatomopathology, biology...).
- Compact Disk (CD) - Imaging from another hospital.
- Online systems: data exchange - PACS (Picture Archiving and Communication System) and all the software solutions; E-mail(s).
- International Commission/Quantitative Analysis reports/Reference Protocols/Multi-disciplinary meeting - The knowledge (from previous experience).
- People/colleagues - The knowledge.

2.1.2 Radiotherapy workflow of one patient

For external radiotherapy, the current workflow for treating a patient can be summarized in the following (not strictly linear) steps:

1. Diagnosis;
2. Multi-disciplinary meeting;
3. External radiotherapy patient consultancy;
4. Planning preparation;
5. Image co-registration;
6. Contouring;
7. Dose prescription;
8. Dosimetry;
9. Treatment;
10. Validating treatment position images;
11. Per-treatment follow-up;
12. Post-treatment follow-up.

External radiotherapy is often complemented with other methods in cancer treatment. Before, during or after the external radiotherapy, there might be chemotherapy, surgery or some other treatments. For instance, commonly for cranial tumors, radiotherapy treatment is recommended to start some weeks after surgery. Those treatments may influence the general workflow of external radiotherapy or change information needed for decision making during radiotherapy treatment planning. Here are examples of some influences:

- In case there has been surgery, additional pre-operative images will be needed;
- In the case of chemotherapy, pre-chemo images might need to be taken into consideration.

Once the patient has been diagnosed and the treatment plan possibilities have been discussed in a multi-disciplinary meeting, and external radiotherapy has been decided as part of the treatment plan, the patient comes to the radiotherapy consultancy. During the consultancy, the process of radiotherapy and the steps involved in it are explained and planned.

The next step is to gather all needed data. For all cases, a planning CT scan is needed, but there may also be other procedures as well (immobilization system, gating training...). Once all information about the patient and the tumor has been gathered, the planning of the treatment can start.

If needed, images are combined in order to get information in a combined way (co-registered). Next, the target volumes with the margins around the tumor and the OARs are contoured on the images. The planned doses and limitations of doses for the tumor and the organs are defined.

The last step, before starting the treatment on the patient, is to create and validate a dose plan that is covering the tumor as prescribed and sparing the OARs as limited.

Now the treatment can start. Treatment is delivered in several fractions over 5-6 weeks unless it is a palliative radiotherapy case. But there can also be hyperfractionated treatments (the total dose of radiation is divided into small doses and treatments are given more than once a day) as well as hypofractionated treatments (the total dose of radiation is divided into large doses and treatment is delivered in few fractions over few days). In most cases, the treatment plan is made prior to the first fraction is used for all the fractions.

Depending on the type of treatment, the treatment position is validated by an oncologist in order to ensure that there is no or limited deviation from the planning position.

In addition, there are weekly follow-up meetings during treatment to evaluate treatment tolerance and immediate secondary effects. After some period of time from the completion of the treatment, there will be post-treatment follow-up meetings to evaluate the outcomes of the treatment. In the following sections, details of the aforementioned steps will be revealed.

2.1.2.1 Diagnosis

The starting point of radiotherapy usually is that the patient has some health complaints and reaches the corresponding doctor. In case of lung problems (difficulties to breath, coughing, etc.), the doctor would be a pulmonologist. In case problems in the head (a headache, dizziness, other neurological problems), the doctor would be a neurologist.

The doctor will conduct series of tests/medical procedures in order to determine what might be the cause for the symptoms. Diagnosis is not always straightforward process - missed or delayed diagnoses of cancer may occur. For instance, negative mammogram may lead to missing breast cancer until the patient returns.

The diagnostics is a complex workflow on its own. For example, Poon et al. (Poon et al. 2012) identified the clinical workflow of breast and colorectal cancer diagnosis consisting of 11 clinical activities and three decision points. If a patient is diagnosed as having cancer, then the specialist refers the patient to the cancer-specialized physicians for the treatment of his/her cancer (e.g., surgeon, oncologist).

A patient is sent to a Cancer Treatment Center after the diagnosis of cancer has been established, and after surgery, if the tumor resection was possible. This means that from radiotherapy workflow point of view the sources of information come from different departments/institutions. This, in turn, means that there may be difficulties in acquiring all the information about the patient that was gathered previously (during diagnosis, pre-/post-operatively).

2.1.2.2 Multi-disciplinary meeting

The multi-disciplinary meeting is a review meeting where all (new) patients are discussed for an optimal treatment plan. Different physicians, such as radiation oncologist, medical oncologist, surgeon and the organ-specific physician are participating in the meeting (Blazeby et al. 2006). The outcome of the meeting should be an optimal treatment care for the patient (surgery if possible, radiotherapy associated with chemotherapy or not, etc.). Not all patients discussed here will have external radiotherapy.

The frequency of these meetings is context-dependent (e.g., country, hospitals). For instance, in France, the multi-disciplinary meeting is roughly once a week. Duration of a meeting is 1-1.5 h and one patient is discussed from a few minutes to 15 minutes, depending on the complexity of the case. Another difference between countries is which types of patients are discussed – all new patients or only complicated cases.

2.1.2.3 External radiotherapy patient consultancy

During the radiotherapy patient consultancy, radiotherapy treatment is discussed and details are decided upon.

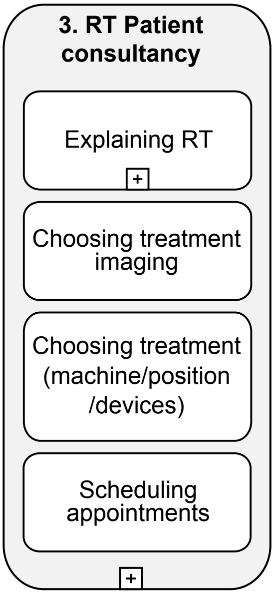


Figure 2-2. RT Patient consultancy

This is the time when the patient (and patient's family) meets the radiation oncologist. The treatment is explained and questions are answered as illustrated in Figure 2-2.

The radiation oncologist has to decide during the meeting what will be the treatment machine, position and what type of custom accessories are needed for patient positioning. Furthermore, the patient is scheduled for the next appointments, such as:

- CT acquisition;
- MRI acquisition;
- PET acquisition;
- Blood tests;
- etc.

After the diagnosis, the patient might have had meetings with other physicians in case the patient has had another type of treatment (chemotherapy, surgery, etc.) also.

2.1.2.4 Planning preparation

The aim of the step "planning preparation" is to gather all data relevant for the treatment. Figure 2-3 gives an overview of "planning preparation". The tasks are not always happening sequentially. For instance, quite often the "patient preparation" is happening immediately prior to the "acquiring planning CT", while the "creating treatment accessories" can happen after the CT image acquisition.

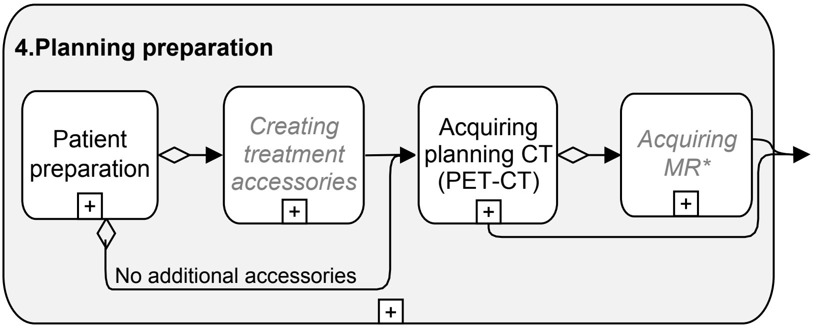


Figure 2-3. Main tasks in planning preparation process

For all the patients there will be tasks of "patient preparation" and "acquiring planning CT" (but not PET-CT). Optionally for some cases, treatment accessories are needed and/or MR images are acquired. Details of these tasks are discussed in the next several sections.

2.1.2.4.1 Patient preparation

The aim of the task "patient preparation" is to prepare everything needed for image acquisition and for patient re-positioning during the delivery of each radiation dose fraction. The main sub-tasks are listed in Table 2-4.

Table 2-4 Sub-tasks of the task "Patient preparation"

Sub-task	**Description**
"Alter" the patient	Make changes in the patient that would support the later parts of the process (e.g., tattoos)
Define patient treatment position	Create and/or define patient positioning devices to be able to reproduce patient's position for each treatment session
Gather patient's files	Ensure all relevant already existing information is available

2.1.2.4.1.1 "Alter" the patient

The aim of this task is to prepare the patient for image acquisitions by specific procedures. For all patients, small tattoos (usually called as BB points or Planning points) are made on the skin. These points will be used during treatment delivery in order to re-positioning patient for each treatment session into the same position as during the planning CT acquisition.

Depending on the type of cancer, the patient might undergo specific preparatory procedures. For example, for prostate cancer it may mean that a procedure is done to implant fiducial gold markers into the prostate. The prostate is known to move a lot between and even during the treatment execution time (Dehnad et al. 2003). Also, the prostate is not too well visible on imaging. Therefore, the gold markers are useful since they are well visible on medical images and based on them, the patient's position can be reproduced better during each treatment session.

2.1.2.4.1.2 Define patient treatment position

The aim of this task is to create the patient positioning devices, and/or define the position of them. The selection of specific devices depends on the type of cancer. For instance, for head or head and neck cases a thermoplastic mask may be created. Depending on the institution it might be created during the CT acquisition time-slot (e.g., if there is a dedicated machine for radiotherapy) or it might be created prior to the CT acquisition, but on the same day (UK).

2.1.2.4.1.3 Gather patient's files

The aim of this task is to ensure that all relevant data about the patient's medical history is available. Patients are not typically diagnosed in the same hospital, or even if they are, it is often not from the same department. Therefore, there are some administrative tasks in gathering all the data of the patient. For instance, pre-operative and post-operative images of patient may exist, but they need to be added to the local software system (used for radiotherapy) also.

2.1.2.4.2 Creating treatment accessories

In some cases, special equipment is needed for the treatment. For instance, for tumors which are close to skin, electrons may be used for the treatment instead of photons. To treat with electrons, an electron applicator will be attached to the head of the treatment machine.

Technician/radiotherapy technologist then have to place an insert into the end of this applicator that is specific to the exact shape and size required for each patient's treatment (2015).

2.1.2.4.3 Acquire planning CT (PET-CT)

CT is acquired for all patients. This is a technological requirement to support the dose calculation for the linear accelerators. For some cases, however, the CT is acquired together with PET, which results in a fused PET-CT image that is also suitable for dosimetry purposes. It may also happen that the PET is acquired independently from CT. That means that the patient's position is different during the acquisition of CT and PET, which will, in turn, require a co-registration step before contouring.

2.1.2.4.4 Acquire MR images

Sometimes there are MRI images acquired (e.g., brain tumor cases). MRI acquisition process consists of a number of steps – T1-weighted MRI before and after injection of contrast enhancing product, T2-weighted, FLAIR (Fluid Attenuated Inversion Recovery), etc. This list of MRI sequences may be even longer in case of a clinical trial – for example, acquisition of MR Spectroscopy Imaging (MRSI) or functional MRI (fMRI). The acquisition of one sequence of MRI takes from 3 to 10 minutes; one full appointment can take up to 1 hour per patient.

2.1.2.5 Image co-registration

Image co-registration (Figure 2-4) is needed in the situation when CT itself is not sufficient for identifying the location of the tumor and other types of image modalities have been acquired.

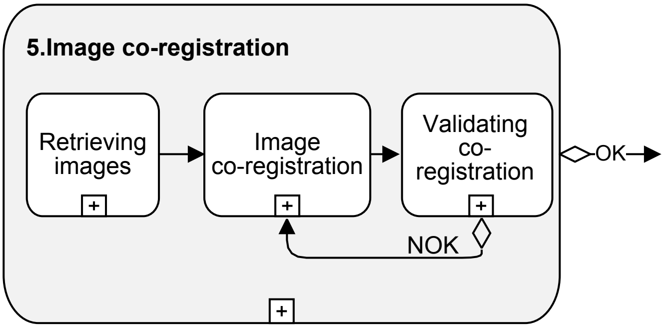


Figure 2-5. Main tasks in image co-registration process

During patient consultancy the radiation oncologist decided on which imaging modalities were needed. In order to use the information from each of the relevant image set, they need to be combined in a good way. There are two types of co-registrations: intra-modality co-registration and inter-modality co-registration.

The goal of "image co-registration" process is to combine and link information from different image sets to allow better comprehension and support the decision making regarding the location of the tumor(s) and OARs. As the planning CT is always needed for dose calculation, typically the co-registration is made with using the planning CT as the reference data set.

2.1.2.5.1 Retrieve images

The aim of this step is to make the images that need to be co-registered available in the system/software. When requests among PACS from different institutions/cities are possible, the images are combined into the local patient folder. Otherwise, Compact Disks (CDs) can be sent.

For specific clinical trial like "SPECTRO-GLIO"(2013a), an on-line database was built for a rapid exchange of imaging and radiotherapy treatment data, and for on-line quality control (only limited access was authorized to specific people of the different centers participating in the trial).

2.1.2.5.2 Co-register images

The aim of this step is to spatially align different image data sets. The co-registration may be performed for the same modality of images (e.g., two T1-weighted MRI acquired on the same patient but at different time = "intra-modality" co-registration) or different modalities of images (e.g., T1-weighted MRI image set can be co-registered to CT scans = "inter-modality" co-registration). The result of the co-registration is a mathematical transformation defined by the "registration matrix". Depending on the software there are possibilities for automatic, manual or semi-automatic co-registration. The exact steps depend on specific software.

The co-registration can be relatively slow process. One of the biggest challenges of co-registration is that most of the times changes occurred between the acquisitions of these two image sets, e.g., the position in which the image has been acquired might have been different, there might have been post-surgical tissue re-organization etc. This produces challenges for the users to decide in which region of interest the co-registration has to be optimal and in which area a shift can be tolerated.

The co-registered images are then evaluated by changing the opacity level and visually evaluating the alignment of some anatomical points (considered as landmarks). Checking of the registration needs to be done in the three orthogonal planes. Axial plane is typically the starting point. Once the images are well co-registered on axial plane, the coronal and sagittal ones will also be checked.

2.1.2.5.3 Validate co-registration

The aim of this step is to decide whether the images are correctly co-registered. Validation of co-registration is happening iteratively. If the images are not well enough co-registered, then there will be changes made in another iteration.

Quite often the co-registration is not done by a senior physician. Therefore, the co-registration has to be approved by a senior physician to be sure that the images are correctly co-registered. If for any reason the images are incorrectly co-registered but still approved, it would mean that the contours will be "translated" to the CT with a spatial shift/error, meaning the spatially incorrectly defined target volume will then be irradiated.

2.1.2.6 Contouring

The assumption for this step is that the correct patient data has been loaded into the system. Contouring process can be simplified into the following (not necessarily linear) steps (Figure 2-6):

1. Delineating the body;
2. Delineating organs at risk (OARs);
3. Delineating gross tumor volume (GTV) --> macroscopic disease;
4. Delineating clinical target volume (CTV) --> microscopic disease, i.e. infiltration;
5. Delineating internal target volume (ITV) --> the expected movement area of CTV during treatment. ITV is drawn only for very few cases (e.g., in lung);
6. Delineating planned target volume (PTV) --> defined by setting some margin to take into account treatment positioning errors.

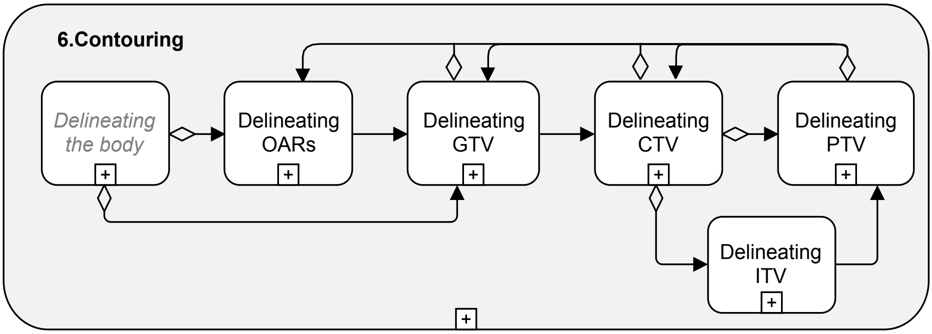


Figure 2-7. Main tasks in contouring process

The order of the above-mentioned steps is typically in this order, but it is not limited to it. For instance, the CTV can be contoured before OARs. Though GTV, CTV, and PTV are dependent on each other, meaning first must be GTV, then CTV and last PTV. Furthermore, for some types of tumors physicians delineate directly CTV without a GTV (e.g., head and neck case) or there is no CTV and only a PTV. Adjustments to contours can be done at any given point – meaning while delineating the CTV, adjustment to the contour of an OAR might be made. Figure 2.1 7 illustrated the spatial relations if the aforementioned volumes in 2D.

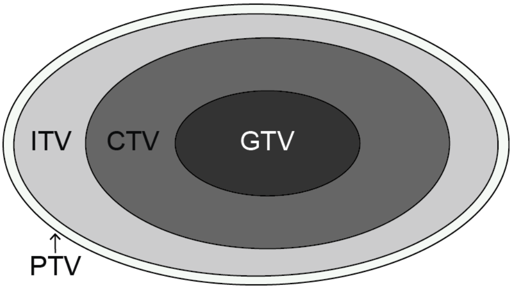


Figure 2-8. Graphical representation of the target volumes, as defined in ICRU Reports No 62 (Parker and Patrocinio 2005).

The general underlying process for any user-dependent contouring is shown in Figure 2-9. A contouring task consists of three main activities "Identify modality", "Identify slice" and "Delineate", and two main decision making points for the questions "Are contours good on this slice?" and "Are there more slices to consider?", respectively.

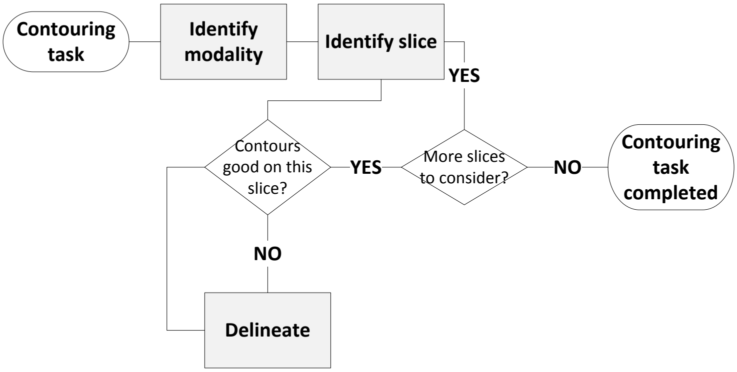


Figure 2-9. The core process of slice based contouring

Each of the tasks and decision making points consists of cognitive and physical tasks as Table 2-5.

Table 2-5 Main cognitive and physical tasks of the core process

Step in the process	**Cognitive tasks**	**Physical tasks**
Identify slice	Decide whether there is enough anatomical and medical knowledge. Compare neighbor slices to decide if the current slice is relevant.	Switch/Scroll through image modality and slice set.
Decision: Are contours good on this slice?	Are there any contours on this slice? Compare contours with neighbor slices. Decide whether the contour captures the full object being contoured. Compare contours on different modalities.	Manipulate the view of the slice (scroll, zoom, and move). Open additional views.
Delineate	Decide where to contour	Contour with the mouse or other tools (basic contouring tools are pencil with adjustable diameter, static or 3D pencil, magic brush...) + Basic object manipulation (Boolean operation, enlargement, reduction, ring definitions)

Decision: More slices to consider?	Is the object visible on the selected slice? Compare with neighbor slices.	Manipulating the view (zoom, move). Switch/Scroll through image modality and slice set.

In the task of defining target volumes within the context of external radiotherapy treatment planning, the key cognitive processes which need support are problem-solving (consisting of information foraging and sensemaking loops) as well as decision making. As part of this view, sensemaking of target volumes can be described as interpreting the medical images and textual reports based on the mental images and models; hypothesis generation of the target volume border location; and evidence finding to evaluate the hypothesis (Aselmaa et al. 2013a).

2.1.2.6.1 Delineating the body

Delineating the body is mandatory for some treatment planning systems (TPSs) since the Source Skin Dose (SSD) is needed to compute the dose inside the volume. Nowadays it is typically done automatically by the software. In addition, the body structure is used to create new structures like “PTV excluding body” in order to crop the part of PTV that is outside the body to avoid air being part of PTV.

2.1.2.6.2 Delineating OARs

One of the axioms of radiotherapy is to maximize the prescribed radiation dose to the tumor while sparing surrounding OAR. This is done by the following two principles – avoiding irradiating unnecessary tissues/organs and reducing the toxicity as much as possible.

Through experience and research, supported by the relatively fast evolution of radiation techniques, quantitative analyses are performed to define recommendations for dose tolerance to organs. These recommendations are regularly updated by, for instance, QUANTEC Steering Committee (Bentzen et al. 2010).

Dose limit is defined per organ and is described as a dose limit for which the organ still preserves its function. There are two types of organs – parallel and serial (analogy to the electricity domain can be made). For a parallel organ, the mean dose for the whole organ matters. For a serial organ, the maximum in any location within the organs matters. This knowledge is important for dosimetry planning.

Table 2-6 Sub-tasks of a task “Delineating OARs”

Sub-task	**Goal of the task**
Identify list of OARs	To know what needs to be contoured according to tumor location
Delineate an OAR	Identify the location of the OAR in the image datasets/modalities
Validate delineations	Validate that the OAR has been properly captured on the image dataset

The delineation process is not always as straightforward as shown in Table 2-6. Depending on the level of expertise of the physician, external knowledge may be needed. For instance, a delineation handbook, guidelines, or automatic atlas segmentation of OARs, integrated into contouring software solutions might be used. Also, discussions with colleagues are common practices.

2.1.2.6.2.1 Identify list of OARs

The first step, as part of delineating OARs, is typically defining (or loading pre-existing) list of OARs into the system (they are called templates). Based on the location of the tumor, it is known what different organs in the proximity are. Quite often this list is adjusted – some organs are removed (For instance, the tumor is located very far from those organs and/or it is known that there will not be critical amount of dose in that area), or some organs can be added following the same principles (e.g. close-by organs). However, with the newer techniques (such as tomotherapy (Mackie et al. 1993)), it has become more important to delineate all OARs. This is because the software automatically optimizes the dose plan and as a consequence may give unnecessary dose to the OARs if priorities are not properly set.

2.1.2.6.2.2 Delineate an OAR

Each of the OARs needs to be delineated. There have been technological advancements which allow automatic or semi-automatic delineation for some organs. For some OARs the existing tools are supporting well the contouring.

For instance, in Eclipse (Varian Medical Systems, Palo Alto, CA), there is a possibility to draw a 3D ball which is good for delineating the eyeballs quickly.

In most of the cases though, the automatic or semi-automatic delineation is not good enough, and some manual corrections are needed after a (semi-)automatic delineation. Unfortunately, if the number of manual adjustments is too high, physicians prefer to do the whole contouring manually. Therefore, the typical process of delineation still has remained manually slice-by-slice contouring on the orthogonal planes supported by interpolation techniques.

Depending on the location of the tumor and the size of it, the list of OARs can become rather long. For instance, for a brain tumor, the list of OARs could be following: Spinal Cord; Brainstem; Chiasm; Optic nerve ipsilateral; Optic nerve contralateral; Eyes; Lenses; Inner ears; Hippocampus; Healthy brain excluding PTV.

The images of the patient are typically viewed in the axial plane, or at least the final evaluation of the contours is done on the axial slices. Such a slice based approach can become extremely time-consuming. A CT scan can be acquired in 1millimeter slice thickness. Roughly this means that to capture 1 centimeter of the organ, contours on 10 slices are needed. In case there is no software support for semi-automatic or automatic contouring, the user has to go through each of the slices manually. Continuing with a 1 cm OAR, the core process depicted in Figure 2-9 would be then iterated at least 10 times.

The software vendors have been working on reducing the time needed for delineation and one of the known functionality is interpolation. The interpolation functionality allows the user to contour on fewer slices. The contours on in-between slices will be interpolated according to the contours of the nearby slices.

Semi-automatic and automatic delineation of OARs are provided to some extent. Unfortunately, in many cases the efforts needed to adjust the automatic contours is not significantly less than fully manual (slice-by-slice) contouring.

2.1.2.6.2.3 Validate delineations

Quite commonly a resident or junior physician does the delineation of OARs in order to gain more experience. Also, since it is a time-consuming task, senior radiation oncologist might not have time for it next to all the other daily responsibilities. Once the contours of OARs have been finalized, a senior radiation oncologist has to validate them. Sometimes this also means that the contours will be adjusted for some OARs, if the senior physician does not fully agree with contours.

2.1.2.6.3 Delineate tumor(s) (GTV)

This step is the core of the overall contouring process. At the same time, this is also the most difficult one. GTV represents the "macroscopic visible" part of the tumor. The "visible" is in quotations because on the images, expansion of tumors is not always that clear. There is a rather high level of uncertainty about where goes the border of the "visible" tumor. Several studies have shown that different physician contour the "visible" border of tumors differently (Weiss and Hess 2003).

Table 2-7 Sub-tasks of the task "Delineate tumors (GTV)"

Sub-task	**Goal**
Retrieve patient information	Get everything needed for understanding the tumor's location /surgery outcome.
Build understanding of tumor(s)'s location	Know where to contour tumor macroscopic expansion
Delineate tumor	Capture the tumor's location on the images
Validate contours	Confirm the location contoured is the appropriate

2.1.2.6.3.1 Retrieve patient information

The common actions for this step might be simply getting the patient's paper folder and/or opening the patient in the treatment planning system. In more complicated cases it might mean that more information is needed from a different institution. For example, if the PET images are not acquired in the same institution – there might be a need to make extra efforts to retrieve the images. As a result of this step, the patient information has been loaded into the treatment planning system – the software solution that is used to do the actual contouring.

2.1.2.6.3.2 Build understanding of tumor's location

The doctor sits down in front of the computer. He/she scrolls trough the images, paying more attention to some slices, less to other. Through this, he/she builds rapidly an understanding of the overall situation (3D mental model) and is now ready to start delineating.

Since the tumor can be anywhere, it takes some efforts to understand where the tumor is. The understanding is built upon past experience and anatomical knowledge (internal knowledge) but also by reading report of the patient and other relevant documents (external information).

Table 2-8 Sub-tasks of the task "Build understanding of tumor's location"

Sub-task	**Goal**
Reading reports of the patient	Attain the clinical picture
Reading patient's history report	Know the disease history
Reading anatomopathological report	Know the description of tumor cells (what type of tumor, how it looked, microscopic spread of tumor etc.)
Reading surgeon's report	Know what was done with the tumor during surgery (partial/complete removal etc.)
Viewing images of the patient	“See” the tumor in the images
Reading treatment related document	Know how the tumor should be contoured

The initial understanding of the tumor is built from reading and viewing, but the understanding deepens throughout the delineation process, while the contours are drawn and adjusted. However, for complex cases, external support is needed to understand the tumor better, for example, radiologists may be consulted.

2.1.2.6.3.3 Delineate tumor

Once the physician has decided where is the tumor (meaning decided where are the “visible” borders of the tumor) the action itself, contouring the border, on the selected slice selected starts. This process is following the core delineation process depicted in Figure 2-9.

2.1.2.6.4 Delineate clinical target volume (CTV)

The medical purpose of CTV is to capture the microscopic spread of the tumor cells, which is confirmed by previous histology studies but cannot be seen on the images (as it is microscopic). Typically it is defined as a margin from the GTV (e.g., for a brain tumor, CTV can be the enlarged GTV by 17mm).

Table 2-9 Sub-tasks of the task "delineate clinical target volume (CTV)"

Sub-task	**Goal**
Generate initial contours	Include medical knowledge of microscopic spread
Adjust contours	Correct contours where it is known medically that the tumor will not be present
Gather relevant information	Get addition external knowledge
Evaluate the contours	Decide whether the contours are good

2.1.2.6.5 Delineate planned target volume (PTV)

The medical purpose of PTV is to capture the possible movement of the organs, patient, and treatment table (setup errors and margins); to be sure that even with those movements the tumor will be properly irradiated. This is done by adding a margin to CTV.

Table 2-10 Sub-tasks of the tasks "delineate planned target volume (PTV)"

Sub-task	**Goal**
Generate initial contours	Include medical knowledge of microscopic spread into the work in an easy way
Adjust contours	Correct contours where it is sure that the tumor will not be present
Gather relevant information	Get addition external knowledge
Evaluate the contours	Decide whether the contours are correct

2.1.2.6.6 Contouring tools

Different Human Computer Interaction (HCI) tools are used in contouring to support different ways of interactions. The common tools are:

- Pencil (also called brush) with adjustable diameter;
- Static or 3D pencil;

- Smart brush (threshold based selection);
- Interpolation (automatically creating the contour based on other contours on different slices);
- “Nudging” based contour adjustments (e.g., Pearl tool in Oncentra Masterplan)

2.1.2.7 Dose prescription

Once the contours are defined, the exact doses for the target volumes and limitations for the OARs will be defined (Figure 2-10).

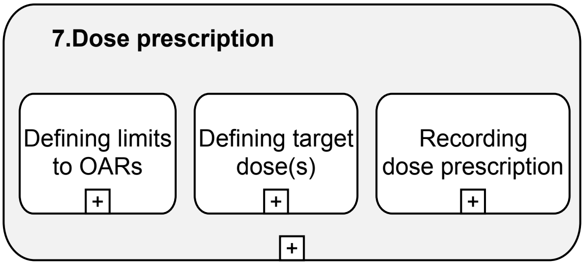


Figure 2-10. Main tasks in dose prescription process

The dose limits to OARs are based on the medical knowledge and available guidelines. There is some level of freedom in how to prescribe the target dose for (planning) target volume. Also, the dose limits may be different when it comes to clinical trials. Below are some examples how the dose can be prescribed.

- Dose to the isocenter/point;
- Minimum dose to 95% of the target volume;
- Dose to the mean of PTV;
- Minimum dose to the target volume;
- Dose to the 95% of isocenter dose.

Last but not least, the prescribed target doses and limits have to be recorded – either as a print, writing on a special form and/or as a voice recording.

2.1.2.8 Dosimetry

The aim of the dosimetry is to create an optimal dose plan that is delivering the prescribed dose to the target volume(s) at the same time respecting the limits defined for OARs.

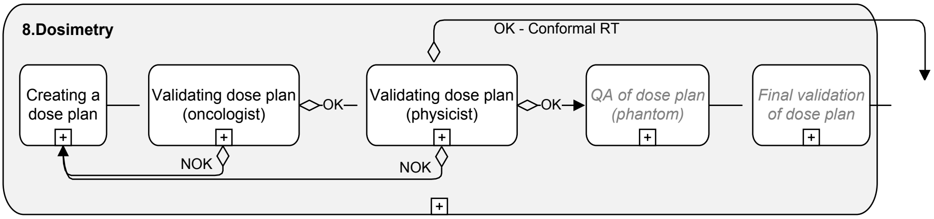


Figure 2-11. Main tasks in dosimetry process

The main steps in dosimetry are:

- Creating a dose plan;
- Validating a dose plan (from oncological point of view);
- Validating dose plan (from physics point of view);
- Quality Assurance of dose plan (on a phantom);
- Final validation of dose plan;

In some hospitals the validation of the dose plan happens during a meeting (might also be called as a multi-disciplinary meeting). A note from a talk with 5th year resident from Germany is:

> *"The dose plans are first validated by a physicist and then at the daily meeting they are discussed."*

2.1.2.8.1 Creating a dose plan

The aim of this step is to create dose plan. The strategies for dose plan creation are different depending on the treatment type. There are also influences on which treatment machine is used because most of them come with dedicated software solutions. During the dose planning, additional volumes may be created (contoured) to aid the dose plan creation.

Direct planning (conformal) is a type of dose planning where the user (dosimetrist) is adjusting beams and wedges in the software to shape the dose delivery. This process may be time-consuming. For example, example times for dose plan creation for 3D conformal RT depending on the location: 30-60 min for head region; 30-60 min for lung cancer; 45-180min for breast cancer. Another type of planning is inverse planning, where the user decides on the weight (priority) of different regions of interest (tumor(s) or OARs) and the software calculates/optimizes the prescribed dose based on them.

2.1.2.8.2 Validating dose plan (oncologist)

The oncologist reviews the dose plan from medical point of view to decide if the dose is covering the target volume and whether the OARs are spared as much as possible from excess irradiation. The main tools for this are:

- a visual review of the dose distribution compared with the contours delineated;
- dose value histogram (DVH);
- isodose contours.

In case the (attending) oncologist is not satisfied with the dose plan created, suggestions are made for changing the dose plan. It can lead to negotiation among different parties regarding what can be done with the dose plan and what the oncologist wishes to achieve.

2.1.2.8.3 Validating dose plan (physics)

The aim of this step is to ensure that the dose plan created is feasible from physics point of view. It also might include checking for accidental mistakes. Depending on the institution and country, the order of validation by oncologist and by medical physicist can be switched – which then also changes the exact tasks within these steps.

2.1.2.8.4 QA of dose plan (on a phantom)

The aim of this step is to deliver the dose plan on the actual treatment machine to a phantom.

2.1.2.8.5 Final validation of dose plan

The aim of this step is to evaluate the dose that was delivered on the phantom is in correspondence with the dose planned.

2.1.2.9 Treatment

The first treatment session (Figure 2-12) of a patient is slightly different than the other daily sessions. There is more attention to different steps as the settings for all the following treatment sessions have to be prepared at that step, therefore, its overall duration is slightly longer. Typically a conventional treatment lasts for 5-6 weeks with 1.8/2 gray (unit of ionizing radiation dose) per fraction. During that time there might be bodily changes to the patient (e.g., weight loss) and also there is an expectation for the tumor to change (e.g., shrinking).

Some patients will have certain prerequisites before their daily treatment session. For example, patient with a prostate cancer should have consistently either a full bladder or an empty bladder (varies between different hospitals) during each treatment session. As such, they might need to drink liquid or empty the bladder before the treatment. During the treatment time, there are at least two technicians in the treatment area: one is focusing on tasks on the computers, the other is positioning the patient on the treatment table and doing the communication and monitors the patient during treatment.

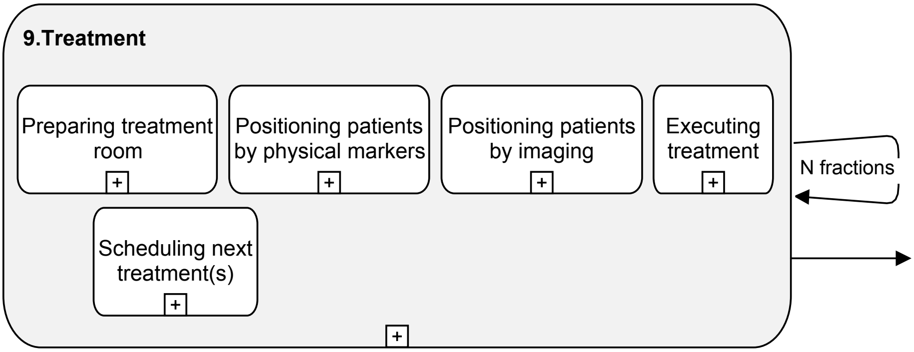


Figure 2-12. Main steps in treatment

2.1.2.9.1 Preparing treatment room

The patient is invited to the changing room while the previous patient is still treated. Patient information is retrieved in the software. After the treatment of the previous patient is finished the room preparation starts. The treatment room is cleaned from the previous patient and the immobilization items are brought out and prepared. Treatment accessories, if needed, are attached to the linac (linear accelerator). Once the room is prepared the patient is invited into the treatment room.

2.1.2.9.2 Positioning patient by physical markers

Each patient has small markers on body (or on the immobilization device). In the treatment room there are laser beams which are used for the positioning of the patient according to these physical markers.

2.1.2.9.3 Positioning patient by imaging

For some type of treatments the patient’s position needs to be verified prior to the dose delivery. This is called image-guided radiotherapy (IGRT). All newer linear accelerators have a possibility to acquire electronic portal images. Onboard imaging can be 2D or 3D. The images can be acquired either as kilovoltage (kV) images or megavoltage (MV) images. In case of 2D images are acquired, they are compared with digitally reconstructed radiographs (DRRs) from the planning CT. In case 3D images are acquired (the cone-beam CT), they are compared with the planning CT. The MV images are of the poorest quality, kV images are better than but still not as good as DRR.

In addition, the image based positioning can happen based on marker match (2D) - For instance, the fiducial gold markers which were placed into prostate during planning preparation are used as indicators for evaluating patient's position compared to the planning position.

In order to be able to compare planning images and the current patient position, the images need to be co-registered. For a simple case, the image co-registration is relatively fast. Sometimes, in case the treatment table cannot be moved automatically, the patient needs to be moved manually and new portal images have to be acquired. With newer machines manual adjustment is no longer needed, the software automatically detects the shift based on the images, and once it is agreed upon by the technician/radiation oncologist, the needed shift of the table is done automatically.

2.1.2.9.4 Executing treatment

Once the position of the patient has been set and agreed upon, the treatment can be delivered. The technician is operating delivery equipment and procedures according to the protocol. Some of the tasks the technicians need to do are selecting the right fields, observing the dose delivery indicators and in-vivo dosimetry indicators. For some types of treatment, the patient is re-positioned and/or additional accessories are added to the linac (repeat of steps 9.1/9.2/9.3).

2.1.2.9.5 Scheduling next appointments

In parallel to other tasks, the next appointments are scheduled in the system.

2.1.2.10 Validating treatment position images

The patient’s position is validated by comparing the planning position and the treatment position. This can be done either directly prior to delivering the treatment or between treatments. In some cases the initial validation of the position is done by the technicians, but the oncologist validates between treatments if the positioning is within acceptable shifts. The oncologist also often validates the position of other oncologist's patients.

One example scenario of validating treatment position images between treatments is as following:

- Read the patient chart;
- Check each on-site image against planning CT;
- Measure distances on planning CT and positioning image – based on a common point to the edge of the PTV drawn on the image;
- The difference that is allowed depends on the margin that is set for the treatment;
- Decisions about the treatment images and whether the position is good:
    - If none is good, they will not be verified and suggestion for repositioning is given;
    - If one is good, that one will be verified, others will be marked as seen.

2.1.2.11 Per-treatment follow-up

The radiation oncologist meets with the patient periodically (e.g., once a week) to discuss the progress of the treatment and to evaluate if there are some side-effects due to the treatment.

2.1.2.12 Post-treatment follow-up

Once the full treatment has been complete, there will be periodical follow-up meetings to evaluate how the success of the treatment was. Example frequencies of first follow-up meetings: every 2 months for glioblastoma, or every 3 months for head and neck.

Prior to the follow-up meeting with a medical doctor, the patient needs to get necessary images acquired (MRI, PET, etc.). Based on these images the physician needs to evaluate the response to the treatment. The patient's disease can be qualified as: Complete Response (to treatment), Partial Response (to treatment), Stable Disease or Progressive Disease according to imaging and clinical criteria. For instance, in the neuro-oncology field, MacDonald criteria (Macdonald et al. 1990) and RANO (Revised Assessment in Neuro-Oncology) criteria (Kaley et al. 2014) are used for classifying treatment responses.

In case there is doubt whether it is a local relapse site, biopsy might be needed and discussion in the multi-disciplinary meeting might happen. The abnormalities seen may be also further metastasis or side-effect from treatment.

During the post-treatment follow-ups, the most important thing is to compare different images acquired before and after the treatment, as well as the radiation dose that was delivered.

Contouring (delineating volumes) as such usually does not happen during follow-up, unless there will be further treatment, it is a clinical research case or there is a need to measure, or it is hospital/national policy/physician's preference.

2.1.3 Conclusions

An overview of the current workflow was given, with a special focus on ‘contouring’. These findings have been presented, discussed and validated by the multidisciplinary partners in the SUMMER project. Nevertheless, they can be different in other countries or institutions.

The 12 main steps identified in radiotherapy workflow are:

1. Diagnosis;
2. Multi-disciplinary meeting;
3. External radiotherapy patient consultancy;
4. Planning preparation;
5. Image co-registration;
6. Contouring;
7. Dose prescription;
8. Dosimetry;
9. Treatment;
10. Validating treatment position images;
11. Per-treatment follow-up;
12. Post-treatment follow-up;

Every step has again many sub-steps; for example, contouring consists of delineating the body, the OARs, the GTV, the CTV, the ITV, and the PTV. In turn, each of these steps again consists of many smaller steps.

In order to (re-)design the User Interface (UI) used in radiotherapy treatment planning, detailed understanding of eye-hand coordination and information processing is needed. This workflow is a first step towards building this understanding.

However, only investigating existing situation might not lead into optimal (and innovative) solutions, as such also new HCI approaches need to be considered (Ramkumar et al. 2013a). Furthermore, novel UIs also require user testing. For instance, in a pilot testing, it was revealed that understanding anatomy in non-orthogonal planes was cognitively demanding for the users (Ramkumar et al. 2013b). As a result of analyzing current situation and existing opportunities from different domain and conducting thorough testing of solutions, the SUMMER project aims at adjusting/changing current workflow of radiotherapy (at a certain level) in a more effective and efficient way.

The workflow overview presented here serves also as a communication aid for the partners and facilitated discussion about projects. The expectation is that it will provide a good basis for the future design work.

2.2 External Radiotherapy Treatment Planning – Situation Today and Perspectives for Tomorrow

The previous section gave a general overview of the radiotherapy workflow. In this section, we focus on the details in the treatment planning phase. The presented research aims to identify the key areas of improvement for software design. In addition, implications of the clinical and technological advancements in the radiotherapy field are discussed. Time-efficiency and effectiveness in radiotherapy treatment planning are identified as the two main areas to be improved. Adaptive radiotherapy, dose painting and proton therapy are highlighted as the main expected drivers of change in the workflow as well as in software solutions.

2.2.1 Introduction

The external beam radiotherapy is an asynchronous teamwork (Munoz et al. 2011). The tasks have been distributed among many team members based on their experience level, expertise, and hospital/national policy. Such distribution allows faster treatment plan creation, but at the same time, this also brings additional risks such as miscommunication.

The technologies used in external radiotherapy are advancing rapidly. More information about the tumors and the OARs, together with extra demands from the radiotherapy team members – pose challenges for physicians to juggle through all this complexity and make the best possible decisions based on that. In order to design a computer-supported solution to aid radiotherapy treatment planning, a deep understanding of the clinical context and the work is required. A solution should fit into the real clinical situation and ensure that it does not cause harm instead of bringing benefits. Human factors and ergonomics focus on the fit between the user, the technology and its carrier, i.e., the equipment and their environments. As such, as the first step towards designing, workflow analysis is found an appropriate approach to understand the medical working environment.

2.2.2 Materials and methods

2.2.2.1 Workflow analysis

Based on ethnographic studies (~40h), naturalistic observations in a radiotherapy department in a French hospital and semi-structured interviews with various radiotherapy team members) workflow analysis were conducted. The main high level tasks were identified, the participants involved were described and time estimates were evaluated.

The first step of workflow analysis was to create a visual representation of the current situation. This workflow diagram has been presented to the same medical staff and also to the project partners and improved iteratively. The next step of the workflow analysis was to conclude analysis of the dominant future trends of external radiotherapy and evaluate the potential changes needed and limitations in the current workflow and also in software solutions from that perspective. The findings were concluded in a summary of areas that need to be adapted in order to support the new ways of radiotherapy.

2.2.2.2 Selective literature review

Workflow analysis was supported by a selective literature review. Scopus, Google Scholar, and PubMed libraries were searched for identifying (1) existing external radiotherapy treatment techniques and (2) current clinical research topics in external radiotherapy treatment.

2.2.3 Results

2.2.3.1 General workflow of one patient treatment

The main phases in the patient treatment workflow currently in clinical practice are following.

- Diagnosis phase– through various activities the patient is diagnosed of cancer;
- Treatment planning phase– through various activities, the information about the patient and the tumor(s) is gathered and the external beam radiotherapy treatment plan is created;
- Treatment phase– the patient retrieves the planned irradiation dose in one or more fractions and the progress is evaluated;
- Post-treatment follow-up phase– the patient has regular meetings with the physician in order to evaluate the response to the treatment.

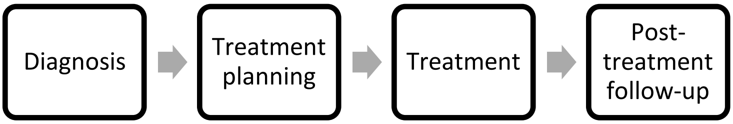


Figure 2-13 Main phases in external radiotherapy treatment

The treatment planning phase, which is one of the focus areas of the SUMMER project, consists of the following events and higher level tasks:

- Multi-disciplinary meeting – a meeting to discuss the overall treatment of the patient (e.g. radiotherapy with chemotherapy, surgery with radiotherapy, etc.). Not all patients discussed at this meeting will receive external radiotherapy. Additional, multi-disciplinary meeting may happen as part of radiotherapy dose plan validation.
- Radiotherapy patient consultancy – a meeting with the patient and their family to discuss the radiotherapy treatment;
- Planning preparation – gathering all the relevant images from the patient's body in the treatment position and any other information needed. Also, patient specific accessories may be created depending on the treatment technique;
- Image co-registration – combination of acquired medical images;
- Contouring – identifying all important volumes such as the tumor(s) and OARs;
- Dose prescription – prescribing the dose for target volumes and limitations to OARs;
- Dosimetry – creating the irradiation plan and ensure the quality of it.

Each of these high-level tasks consists of multiple sub-tasks which are divided between radiotherapy team members (Table 2-11). Each of these sub-tasks in turn also consists of more specific sub-tasks and actions which all need computer-aided support in one form or another. Exact tasks depend on a specific tumor case and treatment technique.

Table 2-11 Main tasks of the phase "treatment planning"

High-level task	**Main possible sub-tasks**	**Possible participants**
Multi-disciplinary meeting	Discuss possible treatments	Radiation oncologist; Medical oncologist; Surgeon; Organ-specific physician;

Radiotherapy treatment planning	Explain radiotherapy	Patient; Patient's family; Radiation oncologist; Secretary;
	Choose treatment planning imaging	
	Choose treatment technique	
	Schedule appointments	
Planning preparation	Patient preparation	Patient; Patient's family; Radiotherapy technologists; Radiation oncologist; Nuclear medicine physician; Radiologist;
	Create treatment accessories	
	Acquire CT (or PET-CT)	
	Acquire MRI	
Image co-registration	Retrieve images	Secretary; Dosimetrist; Radiation oncologist; Medical resident;
	Image co-registration	
	Validate image co-registration	
Contouring	Delineate the body	Medical resident; Radiation oncologist(s); Nuclear medicine physician; Surgeon; Radiologist;
	Delineate OARs	
	Delineate gross tumor volume	
	Delineate clinical target volume	
	Delineate internal target volume	
	Delineate planned target volume	
Dose prescription	Define limits to OARs	Medical resident; Radiation oncologist;
	Define target dose(s)	
	Record dose prescription	
Dosimetry	Create (and adjust) the dose plan	Dosimetrist; Medical physicist(s); Radiation oncologist(s); Technician;
	Validate the dose plan	
	Quality assurance of the dose plan	

2.2.3.2 Current and prospective radiotherapy treatment techniques

There are different treatment techniques of external beam radiotherapy, each of them require slightly different approach. Currently in clinical practice, there are already several different dose delivery techniques used: conventional external beam radiation therapy (2DXRT); 3D conformal radiotherapy (3DCRT); intensity-modulated radiotherapy (IMRT) (Leibel et al. 2002); dynamic IMRT (VMAT by Elekta; RapidArc by Varian;); and more recent techniques stereotactic radiotherapy and stereotactic radiosurgery.

The literature review identified three main topics have been widely researched in radiotherapy field: adaptive radiotherapy, dose painting, and proton therapy. These new concepts of radiotherapy have been recently adopted or are currently researched and prospectively in clinical practice in some years. Each of these requires adjustments to the previously described workflow.

Adaptive radiotherapy is a treatment strategy where the response to the treatment is evaluated periodically between treatment sessions and if needed an adaption is done to the treatment plan. Currently, the concept is researched for various types of tumors (e.g. prostate (Nijkamp et al. 2008), head and neck (Castadot et al. 2010)) in order to indicate positive improvements to the patient outcomes. In order to support this approach, the software solutions need to decrease the overall treatment planning time. Furthermore, different technological solutions for defining volumes are expected for re-planning (e.g. transforming contours from initial planning images to the re-planning images). In addition, there will be more medical images from the patient and tumor movements to take into consideration, and the image co-registration and contouring tasks will become cognitively and technologically more challenging.

Dose-painting technique aims to give heterogeneous dose to the target volume according to additional biological information about the tumor (e.g. for head and neck (Grégoire et al. 2012)). This is achieved either by dose-painting by numbers or dose-painting by sub-volume boost. Both of the approaches are currently clinically researched. Prior to reaching clinical practice, both dose painting techniques require more medical evidence (acquired through clinical trials) about the theragnostic images (e.g. MR spectroscopy) – what extra information do these imaging modalities exactly give and how this extra information can be applied to improve the patient outcomes. In addition, dose painting by numbers requires new algorithmic approaches within the software to support dose prescription (Bentzen 2005). Dose-painting by sub-volume boost requires additional (cognitive) work in order to identify the additional target volumes.

Proton therapy (PT) is a type of treatment where protons are used instead of photons for irradiation. The nature of protons allows targeting the tumor better and reducing dose given on the exit path from the body. Positive patient outcomes have been proven, and more proton therapy centers are being built in recent years (Sisterson 2005). As the main implication to medical practice, there is much higher need to reduce medical uncertainty as the target volumes need to be defined very precisely. Furthermore, the dose plan calculation is different as different physics particles principles are involved.

Each of these new approaches requires some change to the existing workflow. The main demands from the future workflow are reducing the overall time of treatment planning and the higher needs for effectiveness in performing tasks. This can be achieved through technological advancements as well as by supporting the cognitive work of the users.

2.2.3.3 Time-efficiency concerns in treatment planning

One of the main limitations currently is the time taken between the diagnosis and the first treatment session. The main reasons for this are the number of tasks and the time needed to perform each task. As shown previously (Table 2-11), there are many tasks that need to be done by various members of the radiotherapy team. Each of the tasks, consisting of multiple sub-tasks and actions, are dependent on the completion of the previous task, as such the time necessary to complete the planning accumulates into a significant amount. By reducing the time needed per patient without sacrificing the quality of care may bring many benefits such as increased number of patients treated or reduced cost per patient.

Firstly, an image acquisition can take between 30 minutes to one hour depending on the imaging modality, the region of acquisition and if it is done together with treatment accessories. In current clinical practice, CT images are acquired for all patients since it is needed for dose calculations. Furthermore, for soft tissue (such as lung or brain) PET or MRI can be acquired since CT provides limited information about soft tissue. In order to optimize the overall time needed to gather all the relevant images from the patient, there are few options – acquire needed imaging modalities during the same patient meeting (e.g. currently in clinical practice PET-CT acquisition) or overcome the need of CT for dose calculation purposes (e.g. currently researched MRI-based dose calculation (Fotina et al. 2012)).

Secondly, the co-registration of multiple images can be time-consuming. Co-registration aims to combine and link information from different image sets to allow better comprehension and support the decision making regarding the location of the tumor(s) and OARs. Creating a good co-registration can be time-consuming in case the images are not acquired within a small time-frame (e.g. diagnostic CT with planning CT) or when the images are acquired in a different patient position (e.g. diagnostic PET is typically acquired on a soft and more comfortable couch, while planning PET is acquired on a hard and flat surface).

Thirdly, contouring can be very time-consuming. The human body is 3D, but the usual way to display the body on a graphical user interface is by displaying 2D images of the body in different planes (axial, sagittal and coronal) and the contour on these 2D images. There have been advancements in the contouring process such as atlas based automatic segmentation, intelligent software tools such as 3D ball contouring, interpolation between slices, automatic contouring by the software etc. Nevertheless, until now, the tools help up to some level, but there is still significant amount of manual work which is time-consuming.

Fourthly, the dose plan calculation can be time-consuming. Depending on the treatment technique the time-consuming part can be for the user to create the plan (direct planning, e.g. 3DCRT) or for the software to calculate the plan (inverse planning, e.g. IMRT). For instance, for a cranial tumor, the 3D conformal treatment plan creation can take between 30 minutes and 60 minutes, while the IMRT calculation time is 20 minutes. On the other hand, with direct planning one can discuss with colleagues while creating the plan and adjust immediately, while with inverse planning one must wait for the outcome of the dose calculation in order to evaluate if it is good or not.

2.2.3.4 Effectiveness concerns in treatment planning

In addition to the time-efficiency, another concern is effectiveness, the ability to produce the desired results. The effectiveness of treatment planning can be improved significantly in image co-registration and contouring through proper software design and innovative software.

One of the difficulties in image co-registration task comes from knowing which points from different images are the “same” (anatomically). This becomes especially difficult when the patient has been in different positions for the image acquisitions or there have been significant bodily changes – in such cases the cognitive workload for the user increases significantly. One potential (technological) solution for such problems is foreseen in non-rigid co-registration (Crum et al. 2004). Another (technological) difficulty with image co-registration is presenting the multiple fused images in an understandable way as the number of images to take into consideration is increasing with the advancements in the radiotherapy. Currently, the user needs to interpret images in their minds if the number of images is more than two. Furthermore, the medical images are inherently containing uncertainty. Fusion of images is creating additional level of uncertainty which needs to be presented to the user in an understandable way.

Some of the difficulties in contouring are: knowing where is the tumor, where will be the tumor during treatment delivery (e.g. tumor moving due to breathing) and being able to contour the regions of interest effectively. It has been indicated that there can be a high level of inter-observer variability – meaning different expert clinicians identify the tumor volumes differently (Fotina et al. 2012). Furthermore, it has been indicated that even intra-observer variability exists – meaning the same clinician defines the tumor volumes differently the second time (e.g. for soft tissue sarcoma (Roberge et al. 2011)). Higher consistency among different contours is the aim, since it is believed to be more "correct" outcome and as such beneficial for the patient.

2.2.4 Discussion

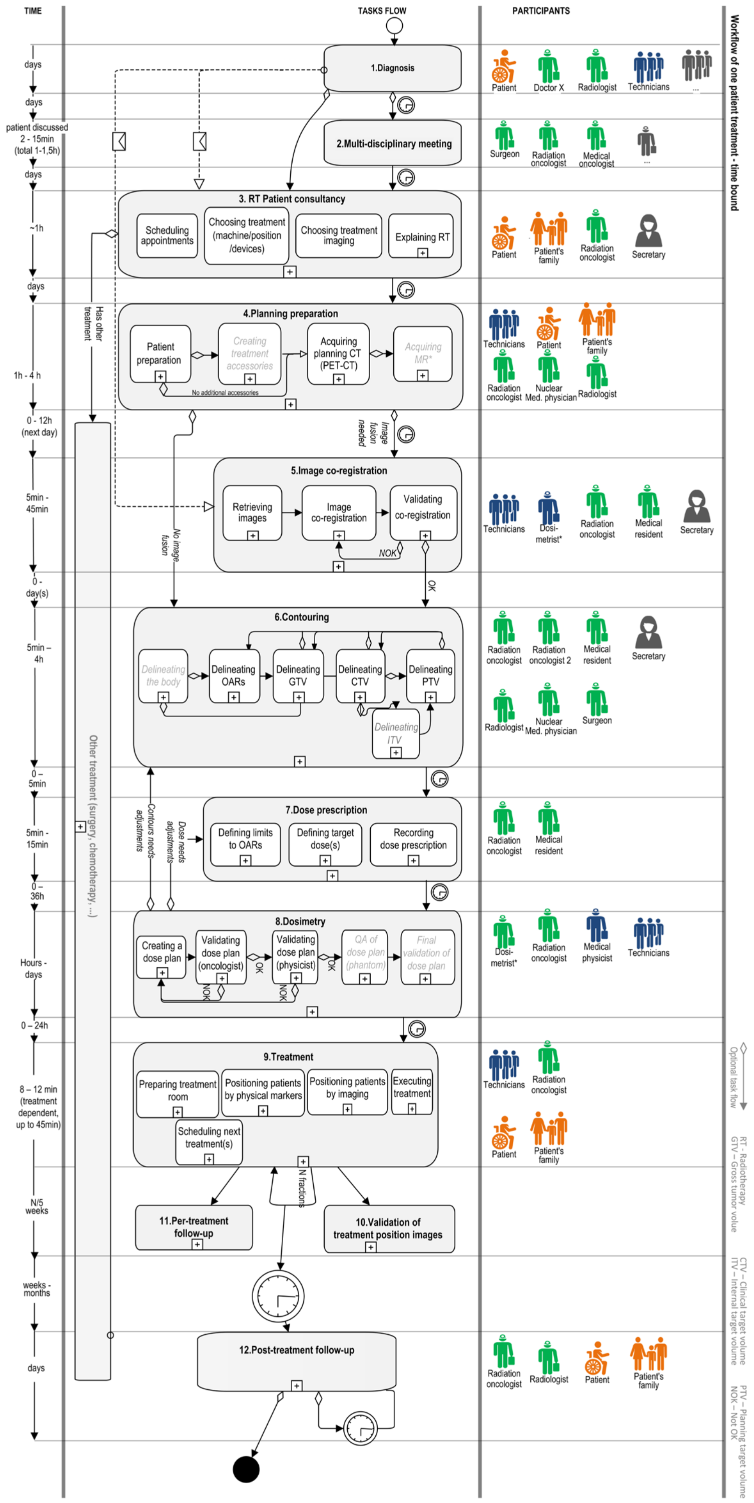
There have been several publications which in one form or another depict the patient treatment (planning) workflow (e.g. adaptive RT (Grégoire et al. 2012), World Health Organization (2008), patient pathway (The Royal College of Radiologists et al. 2008)). The differences between different workflows are based on different focus. This research is mainly based on one French hospital, as such there may be differences compared to other hospitals (in other countries).

A good software system supports user's work by automating user's tasks where possible and supporting the user's cognitive processes in a usable and useful way. In addition, in healthcare IT the graphical user interfaces should reflect the needs of each user (from different disciplines) (Johnson and Turley 2006). For radiotherapy planning software, simply improving the existing solutions might not be sufficient anymore – there are many changes (both increase in medical knowledge as well as technological advancements) and the software solutions need to be receptive to such rapid changes. Furthermore, the increased complexity of treatment planning due to increased amount of information, demands that the software solutions support cognitive tasks in addition to physical tasks.

2.2.5 Conclusion

There are several areas in radiotherapy treatment planning which need improvements from both technological (e.g. better algorithms) as well as from human factors point of view (e.g. reduce cognitive workload, support intuitive interaction). The urgent needs to improve in the radiotherapy treatment planning are to increase the time-efficient of task performance but also the effectiveness. In addition, the task distribution within the radiotherapy team needs to be optimized, though there are already commercial efforts towards facilitating the workflow in radiation oncology department (e.g. workflow tools in ARIA version 11 by Varian).

Identifying the current workflow and envisioning the future directions make it possible to identify what are the main areas to focus on for designing a radiotherapy treatment planning solution which would fit into the clinical context. Furthermore, splitting a complex workflow into smaller tasks allows one to address each of the parts separately while keeping the relations to other areas.


Workflow of one patient treatment - time bound
TIME
TASKS FLOW
PARTICIPANTS
days
days
patient discussed 2 - 15min (total 1-1,5h)
days
~1h
days
1h - 4 h
0 - 12h (next day)
5min - 45min
0 - day(s)
5min – 4h
0 – 5min
5min - 15min
0 – 36h
Hours - days
0 – 24h
8 – 12 min (treatment dependent, up to 45min)
N/5 weeks
weeks - months
days
1.Diagnosis
Patient
Doctor X
Radiologist
Technicians
...
2.Multi-disciplinary meeting
Surgeon
Radiation oncologist
Medical oncologist
...
3. RT Patient consultancy
Scheduling appointments
Choosing treatment (machine/position /devices)
Choosing treatment imaging
Explaining RT
Patient
Patient's family
Radiation oncologist
Secretary
Has other treatment
4.Planning preparation
Patient preparation
Creating treatment accessories
Acquiring planning CT (PET-CT)
Acquiring MR*
No additional accessories
Technicians
Patient
Patient's family
Radiation oncologist
Nuclear Med. physician
Radiologist
Image fusion needed
No image fusion
5.Image co-registration
Retrieving images
Image co-registration
Validating co-registration
NOK
OK
Technicians
Dosi-metrist*
Radiation oncologist
Medical resident
Secretary
6.Contouring
Delineating the body
Delineating OARs
Delineating GTV
Delineating CTV
Delineating PTV
Delineating ITV
Radiation oncologist
Radiation oncologist 2
Medical resident
Secretary
Radiologist
Nuclear Med. physician
Surgeon
Other treatment (surgery, chemotherapy, ...)
Contours needs adjustments
Dose needs adjustments
7.Dose prescription
Defining limits to OARs
Defining target dose(s)
Recording dose prescription
Radiation oncologist
Medical resident
8.Dosimetry
Creating a dose plan
Validating dose plan (oncologist)
Validating dose plan (physicist)
QA of dose plan (phantom)
Final validation of dose plan
OK
NOK
OK
NOK
Dosi-metrist*
Radiation oncologist
Medical physicist
Technicians
9.Treatment
Preparing treatment room
Positioning patients by physical markers
Positioning patients by imaging
Executing treatment
Scheduling next treatment(s)
Technicians
Radiation oncologist
Patient
Patient's family
N fractions
11.Per-treatment follow-up
10.Validation of treatment position images
12.Post-treatment follow-up
Radiation oncologist
Radiologist
Patient
Patient's family
Optional task flow
RT - Radiotherapy
GTV – Gross tumor volume
CTV – Clinical target volume
ITV – Internal target volume
PTV – Planning target volume
NOK – Not OK

Chapter 3

Sensemaking in radiotherapy

This chapter is based on:

Aselmaa A, Goossens RHM, Freudenthal A (2013) What is Sensemaking in the Context of External Radiotherapy Treatment Planning? In: DMD Europe 2013: Design of Medical Devices Conference - Europe Edition 2013. Delft University of Technology, Delft, The Netherlands

Aselmaa A, Goossens RHM, Rowland B, Laprie A, Song Y, Freudenthal A (2014) Medical Factors of Brain Tumor Delineation in Radiotherapy for Software Design. In: Ahram T, Karwowski W, Marek T (eds) 5th International Conference on Applied Human Factors and Ergonomics (AHFE). pp 4865–4875, ISBN: 978-1-4951-2093-0.

Supporting physicians' cognitive processes is one of the future challenges for the software designers. In Section 3.1, some considerations about the cognitive aspects of contouring task are highlighted. In addition, a theoretical lens is provided regarding the cognitive processes involved. In Section 3.2, a study into medical factors influencing physicians' cognition during tumor contouring is presented.

3.1 Sensemaking in the context of external radiotherapy treatment planning

The external beam radiotherapy (RT) is changing rapidly in these years as a result of technological advancements. Despite the expected benefits of integrating new technologies, often it results in increased cognitive workload for the user. This section describes the (1) current clinical context of external radiotherapy from the perspective of defining tumorous volumes; (2) the concepts of problem solving, decision making and sensemaking; and (3) the main cognitive processes while defining tumorous volumes in the frame of user-system-environment interaction.

3.1.1 Background

External radiotherapy (RT) is a type of the medical treatments against cancer, which is to a large extent built on technology – both hardware and software. Although the importance of software solutions is increasing in healthcare (Blumenthal 2009), the current technological solutions are not always fitting to the clinical situations and they also have usability flaws (Chan et al. 2012).

One of the critical tasks for a good treatment plan is to identify the tumor “correctly”. Unfortunately the outcomes depend strongly on the skills of the physician and until now there is no other gold standard (Vorwerk et al. 2009). As a result, for some types of tumors there is large inter-observer variability between experts (expressed by metrics such as volume comparison, center of the volume, concordance index etc. (Jameson et al. 2010)).

In order to precisely identify the tumor, the physician has to build a good understanding of the characteristics of the tumor and the anatomy of the patient, based on medical images, which inherently have a high level of uncertainty. In terms of RT treatment planning, the location and the shape of the tumor is identified by different target volumes (Figure 3-1): macroscopic spread of the tumor as the gross target volume (GTV); microscopic spread of the tumor as the clinical target volume (CTV); the predicted movement of the tumor inside the patient's body during treatment session as the internal target volume (ITV); and the predicted deviation of patient's position during treatment session compared to the planning position as the planning target volume (PTV).

Morphological (CT, MRI) and functional (PET) images – acquired from the patient’s body – are used to identify these different volumes. Even though the technology has advanced significantly in the past decades, the borders/edges of these volumes are still not always clear on these images.

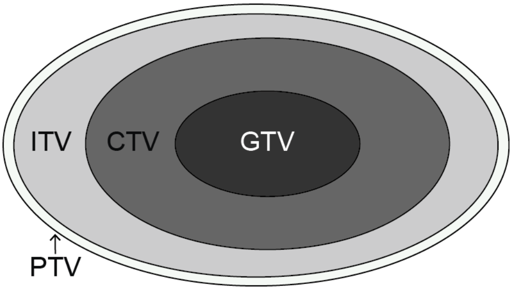


Figure 3-1 Target volumes in relation to one-another in 2D view. GTV – gross tumor volume; CTV – clinical target volume; ITV – internal target volume; PTV – planning target volume

The medical knowledge regarding tumors is constantly growing and new strategies for better treatment planning and dose delivery are researched (e.g., adaptive radiotherapy, "dose painting by numbers" (Bentzen 2005)). This make the situation even more complicated regarding the existing complicated treatment planning, including the process of contouring target volumes. The existing solutions are no longer sufficient to support the RT team in a usable way. The SUMMER project aims to "blend the information in a comprehensible way, and to provide control of multi-modalities in one location" solution (2017a).

The basis of the radiotherapy treatment planning is defining precisely the target volumes and also the relevant organs at risk (OARs). While the organs can be mostly identified based on the anatomical knowledge, identifying the target volumes requires much more cognitive work since there are significantly more variables for the clinician to take into consideration. A well-designed ergonomic software solution is needed in order to decrease the cognitive workload. Such solution should increase the accuracy of the target volume, the user satisfaction, support decision making and consequently improve the patient outcomes. Therefore the main design question is how to support the sensemaking of existing data in order to identify the relevant target volumes through design.

3.1.2 Methods

Ethnographic studies were conducted in combination with workflow analysis in order to identify the context for design. Literature review was conducted to bring in theoretical knowledge from cognitive science.

Ethnographic studies were conducted in the form of naturalistic observations (~40h) in a radiotherapy department of a French hospital and semi-structured interviews were held with various RT team members. The field notes and interviews were then used as an input for workflow analysis.

The first step of workflow analysis was to create a visual representation of the current tasks which was then presented to the medical staff of the same hospital and the project members and improved iteratively. As part of the workflow analysis, hierarchical task analysis (Annett 2003) was conducted, starting with the high level tasks.

The literature review primarily focused on two aspects:

- Problem solving and decision making mostly in clinical context. (Scopus: (TITLE-ABS-KEY("decision making") AND TITLE-ABS-KEY(health OR medical OR medicine OR clinical) returns 200’200+ results) and problem solving (Scopus: (TITLE-ABS-KEY("problem solving") AND TITLE-ABS-KEY(health OR medical OR medicine OR clinical)) returns 24’100+ results)
- Identifying leading theories of sensemaking. (Scopus: TITLE-ABS-KEY(sensemaking OR sense-making) returns 2’300+ results)

The results were explored by publication date, citation count and subject area, based on which the most influential publications were reviewed further. In addition, Google Scholar was used to search for other types of publications (e.g., books, conference papers).

3.1.3 Results

3.1.3.1 Ethnographic studies

From the ethnographic studies the understanding of the working environment was built. Some of the most important aspects related to the task of identifying the target volumes were following:

- At the time of defining the target volumes and also the OARs, all the relevant data about the patient is already gathered;
- The tasks of defining the target volumes and OARs are divided based on the skills required for the specific case (e.g., medical resident or attending physician) and organizational set-up (e.g., technician or physician);
- Validating (checking and accepting) the outcome of each task is of high importance for patient safety management;

- The tasks are performed either individually (initial volumes definitions) or collaboratively (discussion on whether and where the volumes need corrections);
- At any given time there may be interruptions. For instance, radiation oncologists often carry their telephones with them and may be called to treatment room or may be called to consult about another patient;
- Multiple software solutions may be used in order to perform tasks in the most efficient way.

3.1.3.2 Workflow analysis

The RT treatment planning process consists of multiple linear steps, for several of them, it is possible to improve the time-efficiency and the effectiveness (Aselmaa et al. 2013c).

In order to define target volumes and OARs, some of the main cognitive tasks, which in turn will require physical actions, were identified as:

- Building a mental model of the body based on the medical images;
- Processing all information from reports;
- Understanding the visible macroscopic area of the tumor in order to define the GTV;
- Understanding the microscopic spread of the tumor, which is not visible on the medical images but is known from medical research (and the resulting publications) in order to define the CTV;
- Understanding the potential movement of the tumor within the patient's body in order to define the ITV;
- Deciding on the required margins in order to compensate for the treatment positioning uncertainties, in order to define the PTV;
- Identifying the (volumes of) the organs at risk which need to be spared from irradiation as much as possible.

The cognitively difficult part in defining these various volumes is not gathering the needed information, but understanding the relevant parts based on all information gathered.

Currently in clinical practice, volumes are mostly defined by contouring the volume borders on multiple 2D slices from the image set. There are different tools in different software solutions to support this tedious process (e.g., 3D ball, interpolation between slices, atlas-based automatic segmentation) but in clinical practices, they did not demonstrate their ability in sufficiently reducing the time-efficiency to a satisfactory level.

3.1.3.3 Decision making and problem solving

The two most researched cognitive processes in the context of healthcare are decision making and problem solving. In order to support these cognitive processes with a designed solution, a full understanding of them in the design context is needed.

In the view of human as an information processing system, problem solving has been defined as the search in the problem space (consisting of an initial state of knowledge, a set of elements, a set of operators and the total knowledge available) in order to reach the goal state (Newell and Simon 1972). A more general understanding is that problem solving is the process of finding possible solutions. At the same time decision making is about judging the possible solutions and choosing one of them. As such, problem solving typically culminates with decision making.

Problem solving research in healthcare was initially focused on describing the reasoning by expert physicians (Elstein 1978) while decision making research was mainly focused on identifying the deviation from the optimal solution (Ledley and Lusted 1959) by analyzing the reasoning process.

3.1.3.4 Sensemaking

The selective literature review identified the leading theories in sensemaking. Sensemaking is researched in different domains since 1980's, which results in different views and understandings in what is the definition of sensemaking. The most referred theories come from the communication/knowledge management and organizational science.

- Organization science - Weick (1995b) defined sensemaking as "the making of sense" and defined it with seven characteristics: "grounded in identity construction"; "retrospective"; "enactive of sensible environments"; "social"; "ongoing"; "focused on and by extracted cues" and "driven by plausibility rather than accuracy".
- Communication/knowledge management - Dervin (1998) developed Sense-making framework which is built on the assumption that "humans live in a world of gaps: a reality that changes across time and space." Furthermore "the Sense-making metaphor forces us to attend

> the possibility of change [and] this forces our attention to human flexibilities and fluidities as well as their habits and rigidities."

The main theories rooted in the domain of computer science are:

- Decision making/artificial intelligence - Klein et al. (2006b) developed a Data/Frame theory of sensemaking: "frames [stories, maps, etc.] shape and define the relevant data, and data mandate that frames change in non-trivial ways."
- Human-computer interaction – Russell et al. (1993) defined sensemaking as "finding a representation that organizes information to reduce the cost of an operation in an information task. The product of learning loop is the representation and encodon [instantiated schema] set".

3.1.4 Interpretation

Applying the knowledge from cognitive science or any other domain to solve a specific design problem is not a trivial task. In the previous section a brief overview of relevant cognitive theories for the task of contouring target volumes and surrounding organs' volumes has been described.

The identification of the target volumes is an ill-defined problem. Even though the end goal is clear there are several paths to a solution and there can be several different outcomes depending on the problem solver. In contrary, a well-defined problem would have only one solution (e.g., solving a puzzle). Furthermore, the problem-solving task "contouring target volumes" happens at different levels, on individual level as well as on collaborative level while taking into consideration organizational and other existing regulations.

Newell and Simon's (Newell and Simon 1972) model of problem solving, finding a solution strategy by choosing between operators in order to move from one state to another within a problem space, does not encompass the concept of comprehension building. In such a view of problem solving, the comprehension is seen as a preceding process to problem solving and decision making (Patel and Kaufman 2006). Even though this information processing theory is clear when it comes to well-defined problems with one outcome as solution, it is not that obviously with ill-defined problems (Öllinger and Goel 2010).

A wider view on problem solving defines the core activities of complex problem solving as "data ordeals", "wayfinding" and "sensemaking" (Mirel

2004). Similarly problem solving has been described as a combination of “information foraging loops” (processes aimed at seeking information, searching and filtering it (Pirolli and Card 1999)) and “sensemaking loops” to perform a task (Pirolli and Card 2005).

On the contrary of the usual ill-defined problems, in target volumes’ identification the difficulty for the user is not in gathering the right data, but it is in understanding the existing data and making “good” sense out of it. Providing the relevant data in the right way and at the right moment is the biggest design challenge. Previously mentioned theories of sensemaking help the designer to think of different aspects while designing an ergonomic solution, but in their original form they are not easily applicable for such a specific design problem.

Sensemaking as a cognitive process has not been clearly defined – in some views sensemaking and information seeking have been coupled for years, but recent advancements identify that information seeking and sensemaking are separate though interconnected processes (Abraham et al. 2008). Figure 3-2 attempts to position the cognitive process of individual sensemaking of ill-defined problems. Previously mentioned sensemaking theories describe both external as well as internal aspects (e.g., “being retrospective”) of sensemaking. In this section we will focus on the external aspects and the internal aspects will not be covered.

3.1.4.1 User-system-environment interaction from sensemaking perspective

Ergonomics (or human factors) is the scientific discipline concerned with the understanding of the interactions among humans and other elements of a system (International Ergonomics Association (IEA) 2017). In the context of radiotherapy treatment planning, the interaction is between one (or more users) and the software-hardware system. The software-hardware system often consists of multiple software solutions and many of them require a separate set of hardware (e.g., PC, keyboard, mouse).

Usability has become an essential requirement for any product design, and there is room for improvement within healthcare systems (Kjeldskov et al. 2010). One of the reason for usability problems is the mismatch between the designers intent and the user’s goals - the gulfs of execution and evaluation (Norman 1986). Therefore in order to design the system fitting with the user, knowledge is needed on each aspects of use – cognitive aspects as well as physical and environmental aspects.

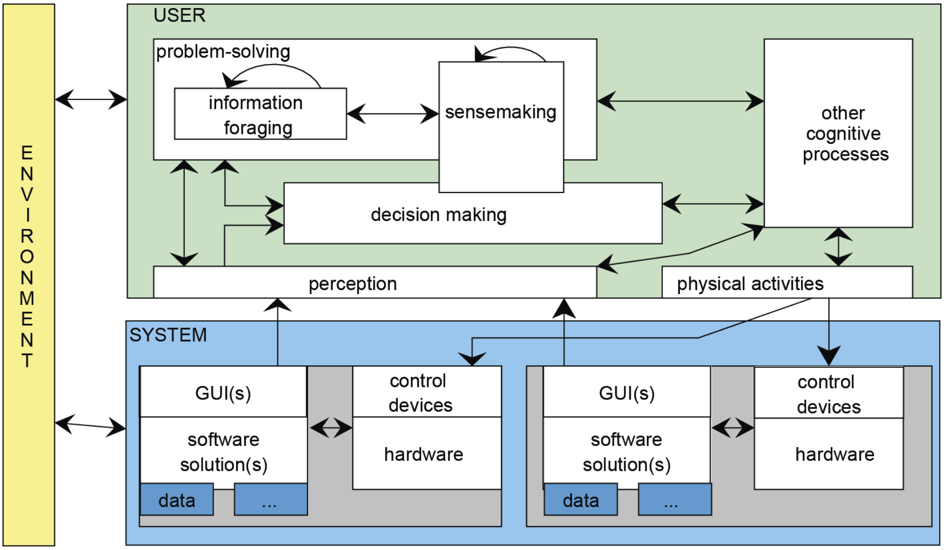


Figure 3-2 External aspects of individual sensemaking

3.1.4.2 *Cognitive processes during contouring*

In the task "contouring target volumes" within the context of external radiotherapy treatment planning, the key cognitive processes which need support are problem solving (consisting of information foraging and sensemaking loops) as well as decision making. For instance, the cognitive processes can be described with the following actions:

- Information foraging – retrieving the images of the patient and manipulating the display of them (e.g., changing contrast level). The main user intent is to have the right information;
- Sensemaking – interpreting the medical images and textual reports based on the mental images and models, hypothesis generation of the target volume border location, evidence finding to evaluate the hypothesis. The main user intent is to understand the information in a right way;
- Decision making – choosing where to contour, deciding if the contours should be adjusted. The course of action can be either to take no action (contours are accepted), look for further information (return to information foraging) or by contouring (matching the contour to the

hypothesis made by physical action). The main user intent is to decide on the right course of action;

As these cognitive processes have different user intents, they also need different design approaches. In order to support the information foraging, best practices and knowledge from information seeking and presentation theories are needed (e.g., the control devices need to support intuitive retrieval of data as well as fast way to manipulate how the data is shown). At the same time, the way the information is shown on the GUI contributes significantly to how sense is made out of the data. Moreover the software-hardware system has to support taking proper intended physical action (e.g., drawing the contours exactly where intended).

The user works in the working environment and as such the surrounding working atmosphere will influence him/her. In addition, quite often there is collaboration happening between colleagues in order to continue the task. As such further investigation is needed on how collaborative sensemaking influences the individual sensemaking and how design can support both in order to achieve the best outcomes during contouring the target volumes and organs at risk.

3.2 Medical Factors of Brain Tumor Delineation in Radiotherapy for Software Design

This section presents research about identifying main medical factors relevant for the delineation in the radiotherapy context. This is seen as a first step in deepening the understanding regarding tumor contouring for software design. Using two discussion formats with six radiation oncologists, 29 medical factors regarding the delineations of tumorous volumes were identified, categorized into: treatment context, tumor context and tumorous areas. In addition, the role of multimodal images, dose planning, as well as future wishes were elaborated.

3.2.1 Introduction

Human and non-technology issues are not trivial to solve in the design of biomedical information systems, since the complexity of tasks and the pressure on users are high (Pantazi et al. 2006). As such, in many healthcare software systems, the interface and interaction design is not always optimal for the intended usage. For instance, in an evaluation of IPLAN® radiotherapy software (Ramkumar et al. 2014), more than 20 usability issues were discovered. Usability problems hamper the effective and efficient usages of the systems, thus further lowering the quality of the service (Teixeira et al. 2012).

Acquiring the clinicians (users)' wishes by asking them is a general practice for improving usability while designing (and developing) a software. By software design one is referring here to the early phase of software engineering process where conceptual interface and interaction design is created, e.g. in the form of prototypes or use cases. However, only asking is not always sufficient since: 1) The perception of the clinicians does not necessarily correlate with the optimal performance (Andre and Wickens 1995); 2) The clinicians' wishes may be limited by their own understanding of the complexity of their work and their design vocabulary. "Give clinicians what they want" without understanding their cognition and actions is insufficient (Karsh et al. 2010). Therefore, to design a better software solution, it is important to study clinicians and their contexts for a better understanding of the complexities of their work - the tasks, processes, contexts, contingencies, and constraints (Karsh et al. 2010).

To understand clinicians and their daily tasks, the Human-centered design (HCD) is often used. HCD was developed for designing interactive systems. It aims at making systems usable and useful by focusing on the users, their needs and requirements by applying human factors/ergonomics, and usability knowledge and techniques (International Organization for Standardization (ISO) 2010). For instance, one of the four main HCD activities named "understand and specify context of use" aims to gather information regarding the users, tasks and organization, technical and environmental characteristics. In the usage of the HCD approach, (Vicente 2010) encourages taking a systems approach in order to solve human factors problems across five different levels in the design: physical, psychological, team, organizational and political level.

In order to design a software solution fitting in the clinical context, a deeper understanding about the medical factors influencing delineation decisions is required. This section presents a research approach for getting a deeper understanding of the psychological (cognitive) level of clinicians regarding the specific task of tumorous volumes delineation.

3.2.2 Radiotherapy

Primary brain tumors are less common (<1.5% of all cancers) in comparison with tumors of other organs (e.g., lung, breast, or colon). However, they are important cases to consider due to the extremely poor prognosis of patients, as well as the histopathologic complexity and biologic behavior of the tumor (Karkavelas and Tascos 2011). Brain tumors are often treated with external radiotherapy in combination with surgery and/or chemotherapy. Radiotherapy damages the DNA of cancer cells. Sufficient doses may prevent the cell from reproducing, or even trigger apoptosis. Although this effect is more pronounced for rapidly reproducing tumor cells, the radiation also damages healthy tissue. Thus it is important to sculpt the dose distribution by targeting the tumor region with maximum dose while sparing the healthy tissue around it as much as possible.

Radiotherapy treatment planning for brain tumors is a complex process involving many participants and tasks. The complexity arises from the fact that radiotherapy must be personalized for each patient. The treatment planning starts with acquiring brain images using a variety of imaging modalities. This is followed by co-registering (fusing) the images to the same coordinate system. A physician then defines the tumorous volumes and organs at risk on those medical images (MRI, CT and/or PET). Once the volumes are defined, the actual dose delivery can be planned. Among different tasks within this workflow, the task identifying the tumor "correctly" would benefit most from an improved understanding of the cognitive processes that are involved in the process (Aselmaa et al. 2013b).

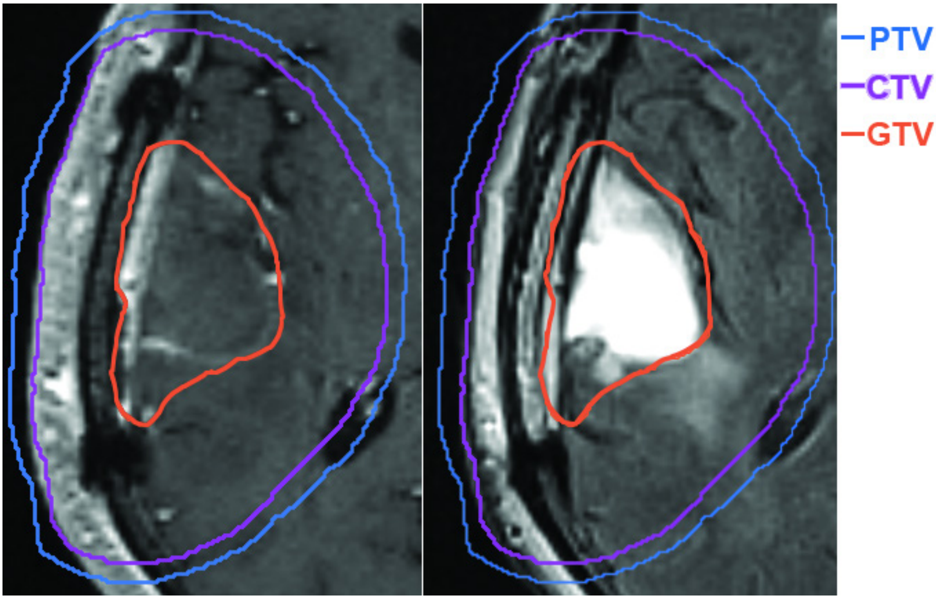


Figure 3-3 Example of "classical" target volumes GTV, CTV, PTV on one axial MR slice of the brain; left) T1-weighted image with gadolinium enhancement; right) FLAIR sequence

Figure 3-3 illustrates different volumes identifying brain tumor. They are: 1) Gross Tumor Volume (GTV) - the macroscopic spread of the tumor; 2) Clinical Target Volume (CTV) – the surrounding estimated microscopic spread of the disease; and 3) Planning Target Volume (PTV), the volume taking into consideration possible positioning errors during treatment delivery. The GTV is delineated on the available imaging modalities based on guidelines, treatment protocols, and the physician's experience. The CTV is then created by adding a margin around the GTV (e.g. for glioblastoma multiforme 2–3 cm (Mason et al. 2007)) and making needed adjustments. Finally, the PTV is created by adding a margin around CTV (for a typical brain treatment it is 3mm).

In the past decades, advances in medical imaging have made it possible to have additional information about the tumor biology and thus know better where additional dose might be needed. For instance hypoxic tumor areas are shown to be more radioresistant (Moeller and Dewhirst 2006). In another example, metabolically active regions detectable with magnetic resonance spectroscopic imaging (MRSI) are predictive for the site of post-RT relapse

(Laprie et al. 2008). In those cases, additional volume(s) might be created (as shown on Figure 3-4) based on which the dose boost area can be defined.

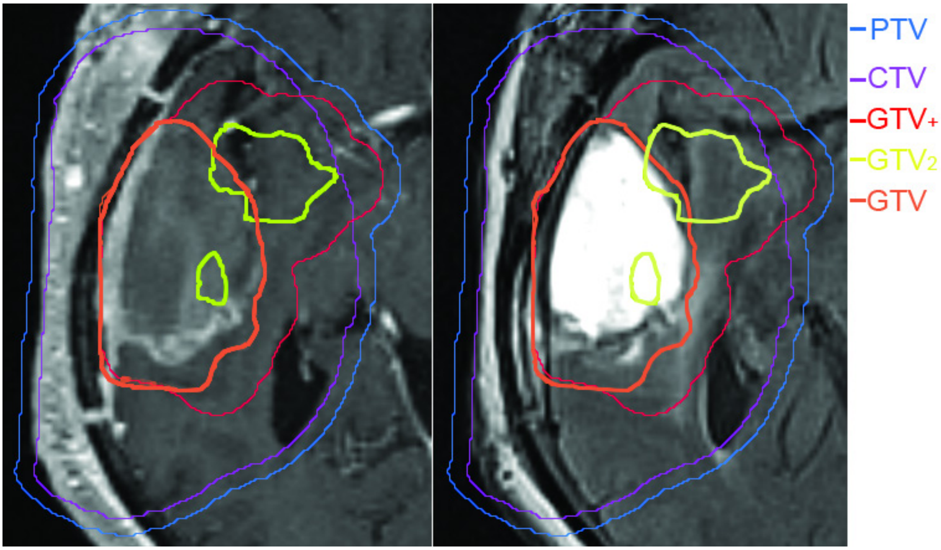


Figure 3-4 Example target volumes for dose boost on one axial slice, GTV2 = metabolic abnormalities, GTV+ = GTV+GTV2+enlargement; left) MRI T1 post gadolinium injection; right) MRI FLAIR

Previous research indicates that the GTV and the CTV are prone to inter- and intra-user variation (Weltens et al. 2001). This is especially critical since the systematic error introduced in the delineation step, will be carried on through the whole treatment planning (Weltens et al. 2001). These differences can be mainly explained by the subjective interpretation of the medical images. In addition, knowledge about tumors is still limited (though growing), resulting in medical uncertainty. Furthermore, users need to deal with a continuously growing number of medical images which makes the task cognitively challenging. All these factors result in variations of defined tumorous volumes.

GTV represents the “visible” tumor. This offers the possibility of automatically detecting GTV by intelligent computational algorithms. This approach may reduce the inter-observer variability (and also the human effort). Research efforts toward this direction can be observed in numerous literatures and conferences. For instance, at the conference Medical Image Computing and Computer Assisted Intervention (MICCAI), there were challenges in 2012 (http://www2.imm.dtu.dk/projects/BRATS2012/) and 2013 (http://martinos.org/qtim/miccai2013/) on Multimodal Brain Tumor Segmentation where multiple algorithms were presented. It can be expected that those automated algorithms will create sound GTV for a subset of cases soon. However, for general clinical use, current algorithms did not reach the satisfactory level. Thus, current radiotherapy software solutions often integrate the technological advancements and the physicians' cognition together as a semi-automated process, and a deeper understanding about the physicians' cognition and medical factors influencing delineation decisions is required in the development of those solutions.

3.2.3 Research approach

In order to gain deep insights into the physicians' cognition, qualitative approach was taken where two parallel approaches: case and list discussion (both consisting of a preparatory task followed by a discussion), were used in order to get a global overview of physicians and their daily tasks. One of the aims while choosing the research approach was not to be limited to physical distances, and to be able to collect information from physicians in several different countries. Due to the busy work rhythm of physicians the aims was to keep the discussions as short as possible (between 10 and 15 minutes per discussion). Another aim was to give preparatory tasks to the participants prior to the discussion in order to trigger reflection on their work and to get them mentally prepared for the discussion. In total, six (female) radiation oncologists with varying levels of experience (3-21 years, starting from residency) from France, Germany and Netherlands participated in the study. Table 3-1 lists each participant and the types of research they joined.

Table 3-1 Overview of participants

	Case discussion	**List discussion**
Participant 1	Yes	No
Participant 2	Yes	Yes
Participant 3	Yes*	No
Participant 4	No	Yes
Participant 5	No	Yes**
Participant 6	Yes	No

** the discussion was more general than one case discussion*

*** no follow-up discussion due to time and language limitations*

For the first approach - case discussion, a preparatory worksheet with open-ended questions was sent to the four participants at least one week prior to the discussion. The participants were encouraged to print out the worksheets and take notes on it. The aim of the case discussion was to discuss in detail one difficult case of a brain tumor they had. The discussions were held over telephone call (three participants) or in person (one participant), and discussions were audio recorded.

For the second approach – list discussion, a preliminary list consisting of 31 medical factors was created. 19 medical factors were derived from the VASARI MR feature list (2013b). VASARI project aimed to develop reproducible methods to classify MRIs of glioma tumors based on the observations familiar to neuroradiologists, in order to describe the morphology of brain tumors on routine contrast-enhanced MRI (2013c). In addition, 12 factors were identified based on prior ethnographic studies (Aselmaa et al. 2013a), workflow analysis (Aselmaa et al. 2013b), lectures given within the SUMMER consortium (2012) and other medical literatures.

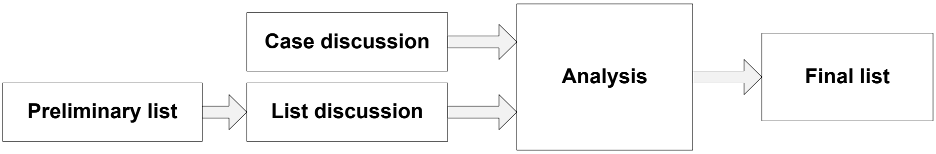


Figure 3-5 Approaches used in the presented research

The preliminary list consisting of 31 medical factors was then formatted into a worksheet, and was sent to the three participants at least one week prior to the discussion. After all participants had filled in the worksheet, a telephone call (audio recorded) was held to elaborate on their responses. With participant #5 there was no follow-up discussion due to a language barrier; however the translated worksheet was included in the analysis.

3.2.4 Results

In this section, we present the identified medical factors (Table 3-2) based on the returned worksheets and the transcribed interviews. The factors are categorized as follows: treatment context (8 factors), tumor context (10 factors) and tumorous areas (11). The last column in the table indicates the number of the participant mentioning the factor. It is important to highlight that if a factor was mentioned by only few participants, it does not imply that other participants did not find it relevant – it only means that it was not mentioned by other participants.

The presented 29 factors all influence to the delineation process to a bigger or smaller extent. However, few of them are either more frequent or have greater impact.

Different institutes have different treatment protocols and approaches to the delineation. It is most evident in the practice of choosing which tumorous area is included in which target volume. Also the margins that are used to generate CTV from GTV vary. However, this is not a new finding. As participant #6 said "the way people delineate gliomas is different in different hospital", and participant #2 "radiotherapy is very different according to the center". Even though there are guidelines, each institute often adapts them to their practice (Chang et al. 2007).

Radiotherapy treatment plan is an outcome of a collaborative teamwork. When there are difficult decisions to make, other colleagues are asked for input. For example, participant #2 mentioned during the case discussion, that "I have asked more experienced oncologist". It was also mentioned by participant #3 "and we also discuss cases". Additionally, collaboration was identified as one of the important factors also during prior ethnographic studies (Aselmaa et al. 2013a). Munoz et al. (2011) investigated the collaboration within the whole radiotherapy workflow and identified that informal communications are the most used strategies to collaborating.

All participants mentioned surgery as a factor. All physicians considered the post-operative (especially with partial resection of the tumor) setting challenging. For example, when they were asked about what was the most difficult part of the case discussed, physician #1 said "how things have moved because there is a surgery" while keeping in mind that "still you don't want to miss the surgical cavity". It was also confirmed by participant #3 that it is challenging to understand what is what: "Well, it's always post-operative setting where it's difficult. […] you have to differentiate between residual tumor after resection and post-operative changes". In addition, movements of tissue also make it difficult to use image fusion as there is loss of information on the fused images. "And the co-registration [fusion] is difficult because you co-register [fuse] and then it's not there anymore"(#1). This is due to the fact that "the anatomy of the brain can be changed after surgery" (#2).

Table 3-2 Medical factor influencing delineation

	Medical factor	**Influence on delineation**	**BY**
TREATMENT CONTEXT	Institute	Radiotherapy is different in each institute. Direct impact on the delineation practices, such as which regions are included in the GTV/CTV or what margins are used for different tumors.	#4 #6
	RT treatment type	The geometrical complexity of contours is limited by technical capabilities of the treatment machine. For IMRT more complex contours are possible. Impact on PTV margins. Stereotactic treatment requires more precision in delineation.	#2 #4 #6
	Re-irradiation	Delineation for re-irradiation is more challenging since it is difficult to separate between residual tumor and the radiation induced changes. The first treatment plan is used for better understanding of the area treated before.	#1 #2 #3 #6
	Curative or palliative	For curative setting there is more effort in trying to take the volume as small as possible. For palliative setting there is less concern for long-term side-effects.	#1 #2 #6
	Strong clinical symptoms	Either increases cautiousness (not to do more damage), or in case of permanent damage would allow including the damaged area.	#2 #4 #5
	Patient had/has chemotherapy	Probability for side effects is greater for combined therapy which means that it can encourage smaller (more cautious) contours. Images prior to chemotherapy might be needed.	#2 #5
	Surgery	More challenging. Tumor bed is to be included in target volumes.	#1..#6
	Imaging modalities	In different cases, different imaging modalities are acquired and available for use. The delineation process is directly influences by the available data.	#1..#6
TUMOR CONTEXT	Type of primary tumor	Defines which imaging modalities to pay more attention to, the expected tumor behavior, which areas to pay more attention to.	#2 #4 #6
	Changes in tumor	The tumor changes between different image acquisitions are taken into account. All sequences should be examined.	#1 #2 #4
	Tumor	CTV margin depends on the histology	#4 #5

	histology		
	Tumor grade	Low grade tumors are less aggressive, thus the CTV margin is smaller. Tumor grade defined which imaging modality is more useful for macroscopic tumor.	#2 #4 #5
	Tumor growth direction	Knowledge about the expected tumor growth direction helps in deciding whether abnormality is tumorous or might be something else	#2
	Close tumor proximity to organ at risk	Extra information (e.g. thin layer MRI or FET-PET) is needed to be able to better differentiate between the tumor and the proximal organ at risk. Adjustments (CTV and PTV) might be made according to the proximal organ at risk.	#2 #3 #5
	Anatomic barriers	Anatomic barriers (e.g. bone) can act as natural border for the tumor.	#6
	Size of the tumor	Very big sized tumors require more attention to reduce the total volume where possible since the harm to the patient cannot be greater than the possible benefit.	#2 #4 #5 #6
	Multifocal or multicentric tumor	Additional locations need to be checked, like spine, for metastasis. Satellite lesions are included in GTV	#2 #4 #5
	Mid-line crossing of the tumor	Midlines crossing contrast enhance tumor is generally included in the GTV.	#2 #4
TUMOROUS AREAS	Tumor bed/surgical cavity	In case the tumor was completely resected, the surgical bed represents the CTV. In the case of partial tumor resection, the surgical bed is included in GTV together with residual macroscopic tumor.	#1 #4 #6
	Macroscopic tumor	The visible tumor; in case of post-surgery referred to as residual tumor	#1.. #6
	Contrast enhancing area	Contrast enhancement on post-gadolinium injected MRI T1 represents the macroscopic tumor. Lack of expected contrast enhancement requires using different imaging modalities.	#2 #4 #5
	Edema	Edema is often included in the target volume. In case the edema crosses midline and is very far from the macroscopic tumor, it might be left out since the whole brain cannot be treated.	#2 #4 #5 #6
	Necrotic	Located within the macroscopic tumor and part of	#2

	tumor	the GTV.	
	Calcification	Included in GTV	#2 #5
	Infiltration	Additional modalities might be needed to understand the extent of infiltration (e.g. CT scan for bone). Infiltrative part of the tumor is included in the GTV. Additional modalities might be needed to understand the extent of the infiltration.	#2 #4 #5
	Satellites	Additional locations need to be checked, like spine, for metastasis. Satellite lesions are included in GTV	#2 #4
	Metabolic abnormalities	Metabolic abnormalities are generally assumed to be tumor and thus included in the GTV. However, the correlations are still being investigated with clinical trials.	#2 #4 #5
	Hypoxia	Hypoxia, one of the metabolic abnormalities, is included in the GTV. However the dose delivered to hypoxic volume might be increased.	#2 #5
	Cyst(s)	Cysts are included in the GTV or CTV, though they are not too common	#2 #4 #5

When the patient has had surgery, the tumor bed (also called surgical cavity or operative bed) is included in the delineated target volumes. Tumor bed typically has contrast enhancement around it. However, whether it should be included in GTV or CTV varies. In the case of a complete tumor resection, the tumor bed is considered CTV as mentioned by participant #4"when there isn't no tumor anymore that's CTV directly". However, if the tumor was only partially resected, normally it means that there is macroscopic tumor - "There is a GTV, yes. When it's partial surgery. Because there is macroscopic tumor"(#4). However, there seem to be differences between institutes and physicians in whether tumor bed is GTV or CTV. As participant #4 mentioned "some radiation oncologists talk about GTV after surgery [complete resection], I think that's false".

One of the common changes in the brain due to the tumor is edema – excess fluid collecting in the intercellular space. Participant #3 mentioned "it is sometimes difficult to differentiate between edema and tumor". Most of the times, edema is considered to be due to the infiltration of the tumor and thus included in the CTV. However, inclusion of edema in target volumes depends on the type of the tumor. For instance, participant #6 gave an example of benign meningioma where the edema is not included. In most cases when edema is present, it encompasses the surgical cavity. Additionally, as participant #2 mentioned, edema in the images may be a result of the prior surgery. In case of such uncertainty, additional medications might be given to the patient to see if the edema will go away, if not then it is assumed to be tumor.

In the current way of defining target volumes (as was shown in Figure 3-3), where the dose is delivered to the PTV, there is no difference to the patient outcomes whether an area is part of the GTV or the CTV since it will be within the irradiation field anyway. However, one of the future visions is to have more complex treatment plans where the heterogeneity of the tumor is matched with complex dose plans. For such treatment approaches, however, it is important to define well all the different areas. This approach is based on the use of theragnostic imaging. Theragnostic imaging refers to the use of information from medical images to determine how to treat individual patients while taking into account scientific progress in molecular and functional imaging, in radiotherapy planning and delivery, and in clinical radiation (Bentzen 2005). Participant #2 mentioned as a future wish that she would like to know “what are the most important places in the tumor anatomy we have to irradiate with larger dose or something like this”.

3.2.4.1 Imaging modalities in delineation process

Delineating tumorous volumes is done on various imaging modalities and image sets, and as such they inevitable influence the process.

3.2.4.1.1 Choosing the right data for the right action

Depending on the institute and the situation, there are different imaging modalities used for different tasks. GTV and CTV are typically delineated on one or more MRI sequences. The common MRI sequences used are T1-weighted (before and after gadolinium injection), T2-weighted and FLAIR. There is always a CT scan for radiotherapy treatment planning due to technical limitations of dose plan calculation. When asked from one of the participants which modalities were used, she said “with mainly T1 gadolinium… and CT scan for planning” (#1).

The variation of whether the newer imaging modalities are acquired depends on the institution and the treatment type (e.g., routine treatment or clinical trial). New MRI imaging techniques evaluate tissue blood flow (perfusion imaging), water motion (diffusion imaging) and brain metabolites (proton magnetic resonance spectroscopy) (Drevelegas and Papanikolaou 2011). “On the diffusion you see the pathways of the tumor growing. And the extent of abnormalities [..] you want to see how bad it is and diffusion helps you. Same with perfusion. And same with spectroscopy” (#1). Additionally PET imaging modality could be used with a suitable tracer (e.g., FET-PET) in order to reduce uncertainties, “FET-PET/CT is a possibility to differentiate between active tumor and pseudo-regression [disappearance of contrast uptake] on MRI” (#2).

Based on the knowledge about the tumor (type, histopathology), physicians know which MR sequence is most suitable for identifying the macroscopic spread. High-grade tumors are expected to have contrast uptake on post-gadolinium injection MRI T1-weighted. The contrast uptake then indicates the macroscopic spread of the tumor. Low-grade tumors, however, do not have contrast uptake, but they do have enhancement on MRI T2-weighted or FLAIR sequences and thus these sequences give a better understanding of the macroscopic spread (#2).

With the advances of radiotherapy, the number of re-irradiations is increasing. However, at the time being there are limited possibilities to differentiate between radiation induced changes and tumor recurrence on the available images. Until now though, there is no one imaging modality which can assist physicians with this question. In one study, 11C-Choline PET/CT was shown to provides an effective mean to distinguish brain tumor recurrence from radiation injury; however 11C-Choline PET/CT also indicated false negatives (Tan et al. 2011).

3.2.4.1.2 Combining information from multiple modalities

One of the challenges with using multiple images is that different images show different information as mentioned by participant #2, “There are a lot of cases where MRI and PET-CT are not the same. They show different volumes”. In such situations physicians need to decide how the overcome the discrepancies. As it was elaborated further “The tumor in the contrast enhancement [in FET-PET] was smaller so I decided to use the PET for delineation [..] as the first imaging. [..] I have done GTV by thresholding for the PET-CT as orientation [..] and then I’ve looked over the MRI and also take in some parts – there was enhancement in [MRI] T2”. The main reason for using multiple modalities is that by combining the information from each of them, the identified volumes are less likely to miss the tumorous cells. This was also brought up during one of the case discussion, “there are parts in the MRI [where] there is no contrast enhancement, but there was uptake in the PET” (#2).

The delineation process consists of continuous checking and adjustment. “When the tumor has not been operated, so there is a GTV [macroscopic tumor]. [Then] I do a margin of 17mm around. Automatically. And then I look on the FLAIR or T2 sequence if all the edema is in the margin. [..] If edema is out the margin I change [..] I enlarge my contour. I consider all the edema is disease” (#4). This was mentioned by all participants that they do a contour (on the modality suitable for the task in hand) and then check on other modalities. In the case of a surgery there is a need of comparison of pre-surgical and post-surgical images. And after a patient had a surgery, there are more images needed to be able to better understand what is what.

MRI T2-weighted or FLAIR is used for delineating edema. “The [MRI] T2 is great to see the edema because we take it in our contour” and “And then I look on the FLAIR or T2 sequence if all the edema is in the margin [..].if edema is out [of] the margin [..] I change my contour” (#4). Additionally PET imaging, if available, could be used as mentioned by participant #3 “In the example of astrocytoma grade II, it is very helpful to have good quality MRI but also to have amino acid PET to differentiate between edema and tumor”.

3.2.4.1.3 Temporal view of images

Typically, there are multiple sets of images of a patient, for example, diagnostic and treatment planning images. However, in case of a surgery there are additional images have been acquired both before and after surgery (depending on the organizational set-up). The practice of using pre-operative images varies between physicians. For instance, participant #6 takes pre-operative images (in the case of glioblastoma) into account only if they have very big difference compared to the planning images. However, different practice is to create GTV by combining what is seen on all the sequences acquired at different moments in time. As participant #2 said “I would try to include all images [in the delineation] of GTV if it is logical and possible”.

One important aspect for delineation is to compare the created contours on the images acquired in different moments of time. “The gadolinium enhancing lesion before surgery and the residual enhancing lesion after surgery” #1. As one physician (#1) was describing “I wanted to take all the place where there was relapse” which is seen in the pre-operative images; “and then all the relapse that was operated and then the little region [of the residual enhancing lesion after surgery]" which are seen in the post-operative images. Another example, mentioned by participant #4"I contour in the post-operative and I verify on the pre-operative". This was later elaborated, "If I contour the tumoral bed, [to verify] I look to the preoperative images. Because we have to.. just look if the tumoral bed is, how to say that, logical.".

Physicians have extensive knowledge about anatomy and tumor biology. As such, they are able to reason the changes seen by comparing temporal images and make a decision on what to include and what not, based on the synthesis of the information. "Sometimes you [..] know how the tumor behaves [..] if I see in this direction [tumor growth direction] something strange I will say it's tumor, but if it's apart from the prior GTV for example then it's maybe another reason" (#2).

3.2.4.2 Dose planning

During delineation, the physicians are estimating the feasibility of covering the target volumes they are creating with the dose they want. As physician #2 mentioned *"If I know that I have the possibility to use IMRT, I probably would tend to draw geometrically complex contours."* This was then further explained, that physicians know about the technical limitations of the geometrical shapes of dose that different types of machines can deliver. Since the aim is to have good dose coverage to the defined target volumes, physicians need to decide already during the delineation what areas to include or exclude in order to reach a geometrical shape that can be sufficiently covered with radiation.

Medical research has shown that tumor is not a homogeneous tissue, but is heterogeneous. Dose-painting represents the idea to visualize tumor subvolumes with a potential resistance problem and to paint some additional dose onto that volume (boost) (Ling et al. 2000). For delineation, however, this means that additional volumes need to be created (as shown in Figure 3-4). For example participant #3 mentioned that she *"would discuss to give a boost on the hypoxic area."* Another situation, where volumes need to be clearly defined, is when there is a dose boost to the macroscopic spread of the tumor, as participant #3 mentioned *"If you give a boost, you would like to boost the tumor, the macroscopic tumor itself, and not the edema"*.

3.2.4.3 Wishes for the future

One of the aims of radiotherapy is to treat the tumorous volume and avoid future relapse in that area. Predicting the tumor relapse area is an ongoing research area. As a future wish, participant #2 said *"if we [could] know the future. In which direction the tumor will grow."* There is constant wish for medical advancement, as participant #3 mentioned *"of course development of new imaging modalities would be great. [..]In the cases when you are not sure of the infiltration and it's the microscopic spread of gliomas would be helpful to have better imaging modalities"*. From technological advancement point of view *"it would be nice to have higher resolution [of medical images] of course."* (#3).

Another difficulty in current work that was mentioned by the participants (#2, #6), is to have the post-surgical images acquired with 48 hours to avoid post-surgical effects, which hamper the readability of the images. However, the radiotherapy department is separate from surgical department or imaging department. Solving this problem would require inter-departmental (institutional) workflow improvements. The participants also mentioned more practical needs for their daily work. Participant #1 mentioned *"I would like to have easy access to diffusion, perfusion and MR spectroscopy data"* and generalized it to *"I would like to have an easy switch from one imaging modality to another"*. Another participant, while emphasizing differences between institutes, mentioned *"It is very helpful if the software is good responding, well responding and fast enough to respond to every kind of input"* (#3).

3.2.5 Discussion

Using the proposed two qualitative research methods, we identified main medical factors influencing the delineation of tumorous volumes. In addition, we also elaborated on the usage of imaging modalities and described the impact of dose planning. Based on these, a deeper understanding of the medical context and the cognitive work of physicians is given. Using the Evidence Based Software Engineering (EBSE) (Kasoju et al. 2013), it is possible to translate those medical factors and the deeper insights regarding the delineation process in to evidences that guide the design and development of software solutions for radiotherapy.

In software (design and) development, requirements play an important role. Using the identified factors as guidelines, it would be possible to ensure that also cognitive and environmental aspects are embedded in the software. For instance, the factor “institute” would require the designer to identify institute dependent practices and to ensure that the designed solution and the accompanying requirements are flexible across institutes. Another factor “imaging modalities” combined with “institute” forces the designer to acknowledge that there are differences in practice on what type of data will be available and as such the designed solution (and requirements) should be able to adapt to the context.

Many of the evidences can be used in the interface design with the ecological interface design (EID) method. EID is a theoretical framework for designing human-computer interfaces for complex sociotechnical systems; its primary aim is to support knowledge workers in adapting to change and novelty (Vicente 2002). As it was seen from the discussion with the physicians, delineating tumorous volumes is a very personalized process. Even though the process might seem to be the same on a higher level, every patient is different and as such the decisions to be made are different.

It became evident from the discussions with physicians, that there are two dominant interactions that physicians’ are engaged in: 1) checking the contour on different medical images; 2) adjusting of the contour, if new information was seen on the different medical images. As such, during software design (and development) special focus is needed in order to reach an ergonomic solution for these interactions.

3.2.6 Conclusions

The aim of the proposed research was to identify medical factors that are influencing the delineation of macroscopic spread of the tumor and surrounding microscopic spread of the tumor in radiotherapy context. Two discussion approaches -preliminary list based discussion and case discussions - were used for exploring the cognitive process. The identified factors were concluded in three categories: treatment context (8 factors), tumor context (10 factors) and tumorous areas (11 factors). More thorough explanations have been given about the impact of surgery and presence of edema. In addition, the role of multimodal images, the relations to dose planning and other interesting findings have been elaborated. These findings can be used as evidences which can support the development of radiotherapy software solutions.

The presented research is the first attempt to bridge the gap between physician’s cognition and software designers. By providing an overview of different medical factors covering various aspects, we expect to support software designers in creating solutions that cognitively fits the task of delineating tumorous volumes. In addition, these medical factors could support software designers in linking clinical reasoning with automatic delineation algorithms.

Chapter 4

Software design

This chapter is based on:

Aselmaa A, Song Y, Goossens RHM (2014) Design challenges in incorporating segmentation methods into radiotherapy software. In: Multimodal imaging towards individualized radiotherapy treatments. p 5, ISBN: 9789461863096

The previous chapters provided detailed insights about the clinical context and physicians' cognitive engagement during contouring. In a technology-driven field such as radiotherapy, existing and upcoming computational advancements, such automation, need to be considered as well. Section 4.1 provides an analysis of the challenges of incorporating multitude of automatic contouring methods into a software design. In Section 4.2, the research software prototype is being described.

4.1 Segmentation methods in software design

Radiotherapy treatment planning is a complex multi-participant process. In a technology-driven context such as radiotherapy, a good software design balances between automation and user interactions. In this section, we discuss the design challenges for incorporating segmentation methods into the radiotherapy treatment planning software, more specifically for the contouring task. Using object-oriented modelling, we identify main design challenges in the categories of general usability, navigation, workflow, and flexibility of interactions. We also highlight that a multidisciplinary approach to the design process is needed to be able to incorporate medical, technical and usability knowledge.

4.1.1 Introduction

Designing software for professionals is a challenge on its own, but designing software in a technology-driven context such as radiotherapy poses even more challenges. On one hand, use of information technology can help decreasing human errors (Pham et al. 2012). On the other hand, poor usability can severely hinder the effectiveness of clinician's work (Viitanen et al. 2011). As such, it is necessary for the software designer to become familiar with the medical needs, working environment as well as with the technological advancements. Only based on a good understanding of the above knowledge, it is possible to propose an initial design concept that could be further improved through co-design sessions.

Radiotherapy is a complex, multi-participant process (Aselmaa et al. 2013c). The full treatment planning process involves multiple clinicians and can take from hours to days to be completed. Contouring, one of the sub-processes of the treatment planning where the contours of all important regions of interest (ROIs) are created, has been identified as the weakest link in the treatment planning (Njeh 2008).

The contouring process begins with defining the list of ROIs to be contoured. This is then followed by contouring each of these ROIs (as depicted on Figure 4-1). Most ROIs are independent from each other and can be contoured in any order. However, some ROIs (e.g., GTV and CTV) are dependent on each other and need to be contoured sequentially. In addition, in clinical practice, the initial contours are often created by a resident, and therefore a more senior oncologist needs to validate (and adapt if needed) the contours.

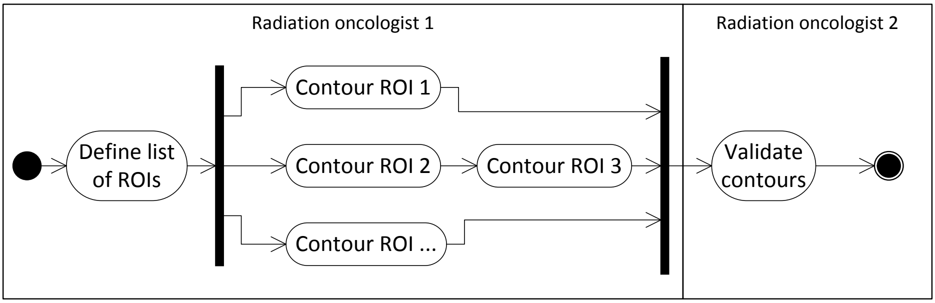


Figure 4-1 Simplified activity diagram of a contouring process

The process of contouring a ROI depends on the types of ROI and the software used. The types of a ROI define which image modality should be used. For instance, skull can be clearly visible on a CT scan. The software used, however, defines which types of segmentation methods are available (as shown on Figure 4-2).

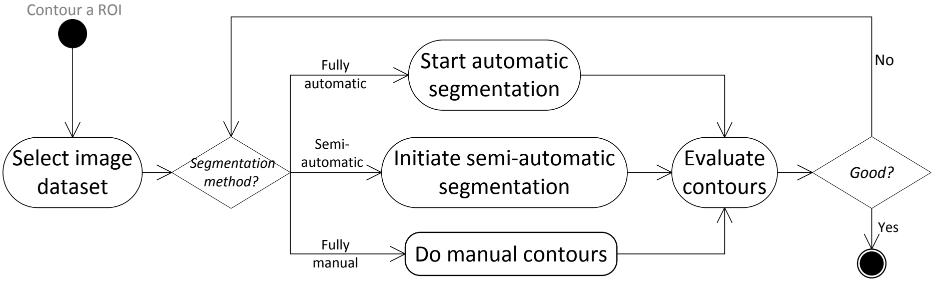


Figure 4-2 Simplified contouring process of a ROI for three types of segmentation methods: fully automatic, semi-automatic and fully manual.

A segmentation method is a specific tool or an algorithm that enables the user to segment (contour) a ROI. Within this thesis, we classify segmentation methods into three categories: fully automatic, semi-automatic and fully manual.

- A fully automatic segmentation method requires no input or interaction from the user for creating contours (except starting it).
- A semi-automatic segmentation method combines computational algorithms and user interactions in the creation of contour. The algorithmic support can vary from seamless to the user (e.g. 3D "Smart Brush" (Parascandolo et al. 2013)) to almost fully automatic (e.g. user input is only required for initialization).

- A fully manual segmentation method assumes no extra algorithmic support from the software (e.g. the line is drawn exactly following the movement of the mouse cursor).

Automation has a lot of potential for many tasks in radiotherapy treatment planning. For instance, the development of (fully or semi-) automated image segmentation methods is one of the key topic of research (e.g., the Brain Tumor Segmentation Challenges (BraTS) at MICCAI (Medical Image Computing and Computer Assisted Intervention) conferences in 2012 – 2014). Current methods of automated segmentation are usable in certain situations; however, it would be necessary to define which methods are usable for which ROIs, and on which types of datasets (Whitfield et al. 2013).

As in any software interface design, general usability principles need to be taken into account. For example, Nielsen (Nielsen 2012) defined the main quality components of usability as learnability, efficiency, memorability, errors, satisfaction, and utility. With the increasing number of segmentation methods available, designing a software interface that is able to balance the automation and the user interaction with high usability will be a challenge.

The aim of this section is to discuss possible scenarios of using segmentation methods for contouring regions of interest, and to highlight different design challenges posed by those use scenarios. For identifying these scenarios, an object-oriented approach was taken. And then based on the identified use scenarios, the design challenges were summarized.

4.1.2 Object-oriented view on contouring

Object-oriented modelling approach allows describing relevant objects and the relations among them in a compact way. It allows identifying different use scenarios, which then can be used as a basis for the interface design process. In this research, the Unified Modeling Language (UML) object diagram was used for modelling objects and relations involved in the contouring process. Typically, UML diagrams are used in software engineering. However, the use of UML diagrams is not restricted to this area and there is increased interest in using UML diagrams for describing other higher level (e.g. business (Eriksson and Penker 2000)) processes.

In a high-level overview of contouring process, the main objects involved are ‘tumor’, ‘patient’, ‘ROI’, ‘image dataset’, ‘segmentation method’, and ‘user interaction’ (Figure 4-3). The main relations between any pair of these objects can be summarized as follows:

- The list of ROIs depends on the tumor and the patient;
- A ROI is identifiable in one or more image datasets;
- A ROI has one or more segmentation methods suitable for segmenting it;
- A segmentation method uses one or more image datasets;
- A segmentation method can have no user interaction or numerous user interactions;
- A segmentation method can be able to segment one ROI or multiple ROIs.

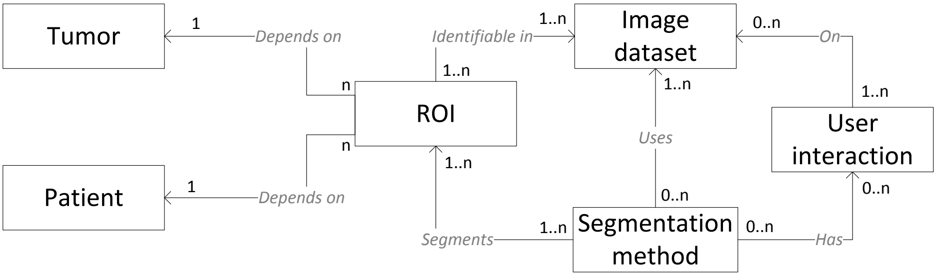


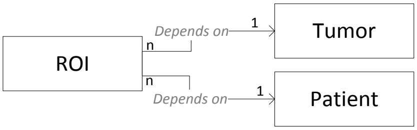
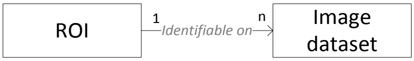
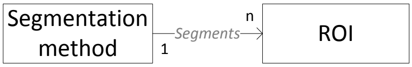
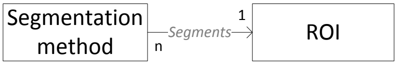
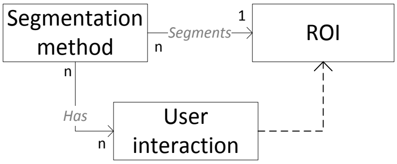
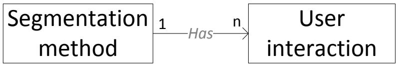
Figure 4-3 Simplified object diagram representing all potential relations between main objects within ROI contouring process

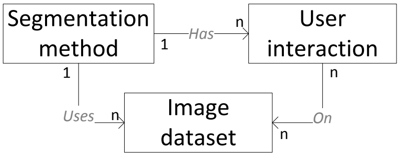
As a practical example, edema is one of the ROIs in contouring a brain tumor. Edema can be identified well on MRI T2-weighted images or MRI FLAIR images (Aselmaa et al. 2014). A fully automatic brain tumor segmentation method called ABTS, is claiming high success rate in segmenting edema present for glioblastoma multiforme (GBM) by using MRI T2 and MRI FLAIR image datasets (Diaz et al. 2013). In addition, their segmentation method is also able to segment another ROI - the GTV.

For most ROIs, the best suitable image dataset for contouring is known from clinical practice and medical research. At the same time, there is a growing knowledge on which segmentation method performs well for which types of ROI(s). Therefore, it is feasible to find an optimal segmentation method(s) for specific ROIs within a software solution.

Designing such a software system that is incorporating multiple segmentation methods for multiple ROIs is not trivial as there are various realistic scenarios of creating ROIs by using different segmentation methods. Each of these scenarios gives additional design consideration. One-to-one relations (e.g. a ROI is identifiable only on one image dataset) on their own do not pose design challenges compared to one-to-many scenarios. However, to enable all different scenarios in the design of a single software solution poses usability and interaction design challenges. Table 4-1 summarizes the design challenges posed by one-to-many use scenarios

Table 4-1 Design challenges posed by one-to-many use scenarios

Main one-to-many use scenarios		**Design challenges**
	There are multiple ROIs for the tumor of one patient	- Navigation between ROIs - Managing segmentations of dependent ROIs
	A ROI is identifiable on more than one image dataset	- Intuitive navigation between image datasets
	A segmentation method segments multiple ROIs	- Intuitive use within workflow
	One ROI has more than one suitable segmentation methods	- Balance between user freedom and cognitive load - Navigation between different segmentation results of a ROI - Creation of a composite contour based on multiple segmentations
	... and the segmentation methods require different types of user interaction	- Consistent user interactions - Intuitive use within workflow
	A segmentation method requires substantial user involvement	- Clear user interactions - Minimized amount of interactions - Balanced interactions

	... and uses more than one image datasets	- Intuitive navigation between image datasets

4.1.3 Design challenges

Identifying different scenarios enables designing a solution fit the clinical needs better. However, the priorities of these scenarios will depend on the number and types of segmentation methods incorporated into the software. Implementing too many segmentation methods can become costly without bringing significant benefits. At the same time, not having enough segmentation methods will hinder the usability (e.g. fully manual segmentation methods require too much time from the users and thus it does not support efficiency).

A starting point for such a software design is to review available segmentation methods (fully automatic, semi-automatic and fully manual) and for each of them, to specify the image dataset types suitable as the input(s), their success rates for different ROIs, and also the required user interactions. For example, Zhu (Zhu 2013) investigated user interactions for three semi-automatic segmentation methods (parametric active contours, geometric active contours and graphical models) and proposed optimal user interactions for them. However, their work did not present an overview of the success rate of segmentation methods for segmenting specific ROIs based on specific image datasets.

Once there is a sufficient knowledge base available to incorporate segmentation methods, detailed graphical user interface design work can begin. In this design phase, the design challenges we have identified in the previous section need to be tackled. Table 4-1 presents those design challenges which were categorized into four categories (Table 4-2): general usability, navigation, workflow, and flexibility of interactions.

Table 4-2 Summary of the main design challenges to be addressed within the design of a software incorporating numerous segmentation methods

Category	**Design challenge**
General	Minimized amount of interactions
	Clear user interactions

usability	Consistent user interactions
Navigation	Intuitive navigation between image datasets
	Intuitive navigation between ROIs
	Navigation between different segmentation results of a ROI
Workflow	Managing segmentations of dependent ROIs
	Intuitive use of a segmentation method within the workflow
Flexibility of interactions	Creation of a composite contour based on multiple segmentations
	Balanced interactions
	Balance between user freedom and cognitive load

4.1.4 Discussion

We have highlighted several design challenges based on different envisioned scenarios. However, the real challenge will remain in reaching a user interface design solution that solves all of these in an effective manner. Poor communication among different stakeholders has been highlighted as one of the main reasons for the failure of software projects (Charette 2005). Even though we have used object-oriented approach for identified different use scenarios of segmentation methods, we consider that it is necessary the interface design process itself follows user-centered design approach (International Organization for Standardization (ISO) 2010).

The ISO standard 9241 (part 210) highlights the needs for a multidisciplinary design team and an iterative approach in software development. For the design of a software solution incorporating numerous segmentation methods, tight collaboration between developers and users is a prerequisite for the success of the development.

4.1.5 Conclusion

In this section, based on object-oriented modelling, we have highlighted different use scenarios with segmentation methods and discussed the design challenges in incorporating numerous segmentation methods into a single software solution. Those various design challenges are categorized into four categories: general usability, navigation, workflow and flexibility of interactions. To tackle those challenges, a multidisciplinary design team, which is able to incorporate medical, technical and usability knowledge, is often needed.

The next step of this research would be to generate possible interface design prototypes to tackle these challenges. Ideally prototypes would be improved iteratively in collaboration with clinicians and developers of segmentation algorithms. The feasibility of this concept will need to be evaluated.

4.2 Research software prototype description

The research approach of this thesis required a software prototype that would simulate the contouring task environment. In this section, the developed software prototype and the variations of it are presented.

4.2.1 Background

In the development of the software prototype, collaboration was established with Netherlands Cancer Institute (NKI), and their contouring research software (described by (Steenbakkers et al. 2005)) was used as a starting point (see Figure 4-4 top) in the development of the research software prototype.

In this thesis, the research software is being also being referred to as (software) prototype since it was adapted for specific (limited) list of functionalities and for pre-defined data.

4.2.2 Technical specifications

The prototype was developed with Delphi and C++. The software has been previously used with different types of input devices (Multi-Institutional Target Delineation in Oncology Group 2011). For the purpose of the current research, mouse and keyboard were used as the input devices.

The modifications to the software code were done on the same computer (laptop) that was used during the studies (Dell Precision M 4700, Windows 7 Enterprise 64-bit, 16GB RAM, Intel Core i7-3720 @ 2.60GHz) eliminating the probability of unexpected performance issues. During the study, the screen of the laptop was mirrored to a larger monitor, and an external mouse and a keyboard were connected.

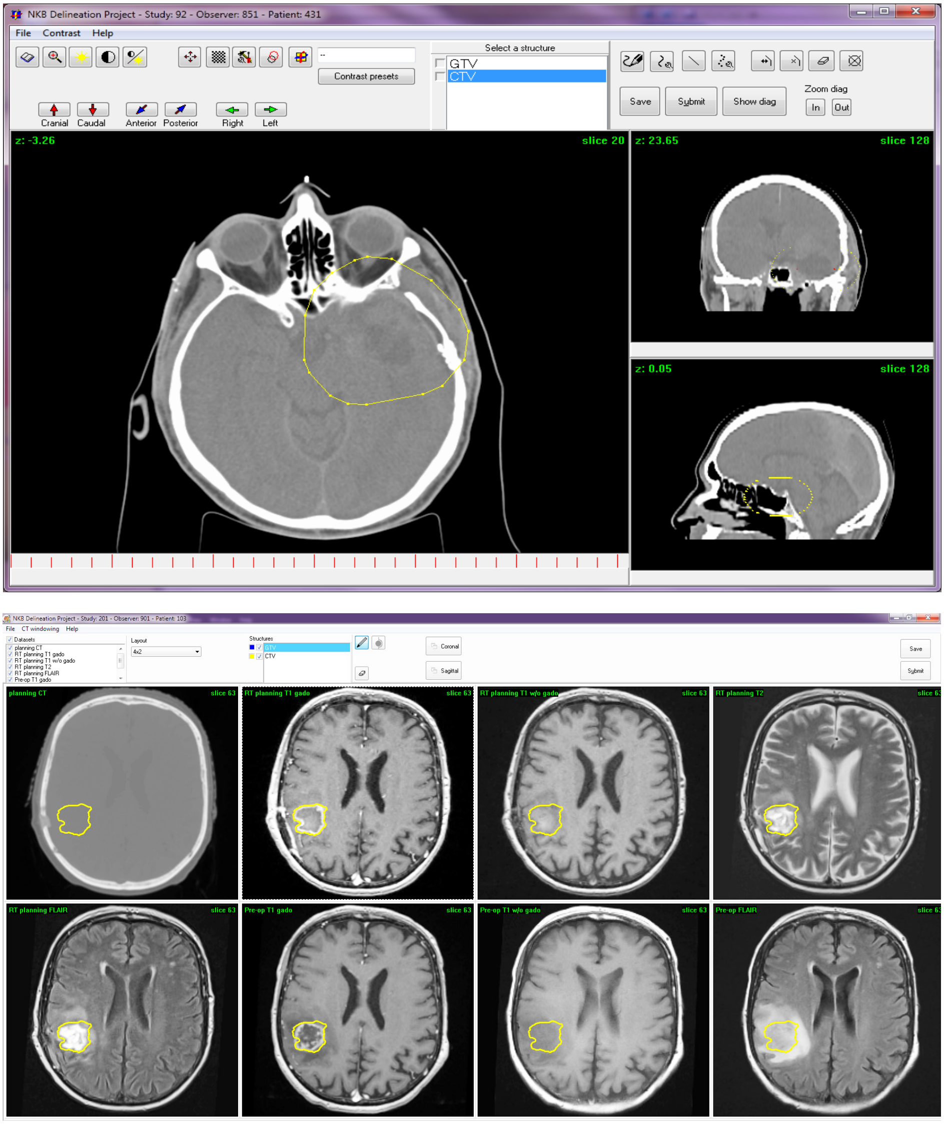


Figure 4-4 A screenshot of (top) the original software prototype, (bottom) the modified software prototype

4.2.3 Modifications to the original software

The research focused on a type of brain tumor - Glioblastoma Multiforme (GBM). Modifications on the existing software were required to accommodate the needs of the research. Analysis of the existing software functionalities was conducted, and the relevant data objects were identified as Figure 4-5.

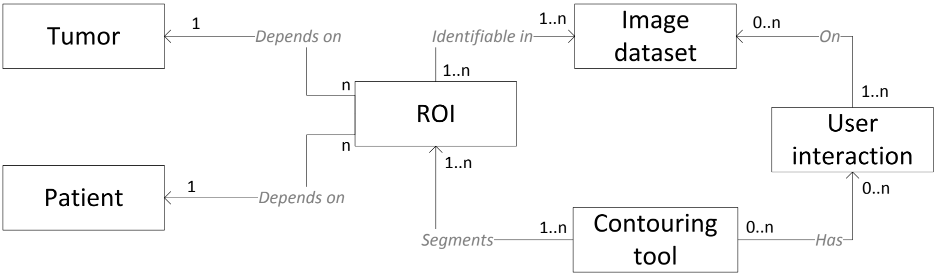


Figure 4-5 The descriptive model of main objects of the software prototype

An overview of the details of each object is listed in Table 4-3.

Table 4-3 Descriptions of the objects within this

Object	**Prototype**
Tumor	GBM
Patient cases	3 for the studies + 1 test case
ROI	GTV; CTV; Contrast Enhancement (CE) tumor; Surgical cavity; Edema
Image dataset	Eight datasets of a patient: pre-surgery MRI T1; pre-surgery MRI T1-CE; pre-surgery MRI FLAIR; RT planning CT; RT planning MRI T1; RT planning MRI T1-CE; RT planning MRI T2-weighted; RT planning MRI FLAIR
Contouring tool	Polygon drawing; Freehand drawing; Nudge; Between slice interpolation; Freehand deleting; Within slice contour deleting
User interaction	Button click; Key press; Right/Left mouse click; Mouse click-and-drag; Changing nudge radius

4.2.3.1 Image datasets

During the contouring of GBM, different imaging modalities are required for viewing and contouring different types of ROI. Based on the medical literature (e.g., (Drevelegas and Papanikolaou 2011)) and ethnographic studies ((Aselmaa et al. 2013b)), eight datasets were identified to be relevant and obtainable for the study: (1) pre-surgery MRI T1-weighted (MRI T1); (2) pre-surgery MRI T1-weighted Contrast Enhanced (MRI T1-CE); (3) pre-surgery MRI FLAIR; (4) radiotherapy treatment planning CT; (5) radiotherapy treatment planning MRI T1; (6) radiotherapy treatment planning MRI T1-CE; (7) radiotherapy treatment planning MRI T2-weighted and (8) radiotherapy treatment planning MRI FLAIR. Data regarding three patients were gathered and integrated into the prototype. In addition, one dataset was added for training purposes.

4.2.3.2 Task workflow with the software

The general expected workflow with the software is shown Figure 4-6. In addition, it was possible to save the interim work at any moment and continue afterwards.

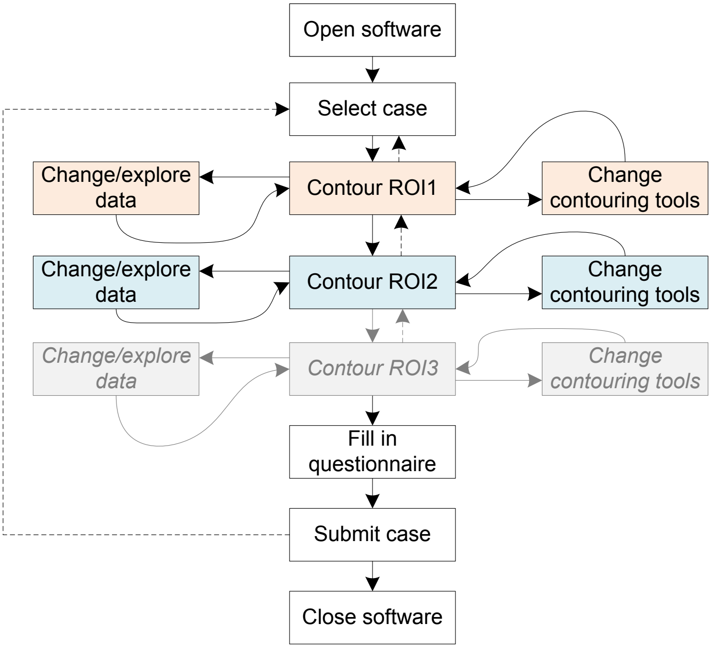


Figure 4-6 General workflow with the software prototype

Within the prototype, two workflows settings were integrated: traditional contouring workflow (ROIs were GTV, and CTV), and sub-region based contouring workflow (ROIs were surgical cavity, CE tumor, and Edema). Each of these workflows had a workflow-specific functionality in the prototype:

1. Traditional workflow incorporated automatic CTV generation as a 2 cm uniform 3D expansion from the GTV.
    a. When starting the software, the planning CT, and MRI T1-CE were visible.

2. Sub-region based workflow incorporated default data display layouts for each structure.
    a. When starting the software, the planning CT, and MRI T1-CE were visible;
    b. Surgical cavity – the planning CT, and MRI T1-CE were immediately visible;
    c. CE residual tumor – the planning MRI T1 T1-CE, and MRI T1 were visible;
    d. Edema (post-surgery) – the planning MRI T2 and MRI FLAIR were visible.

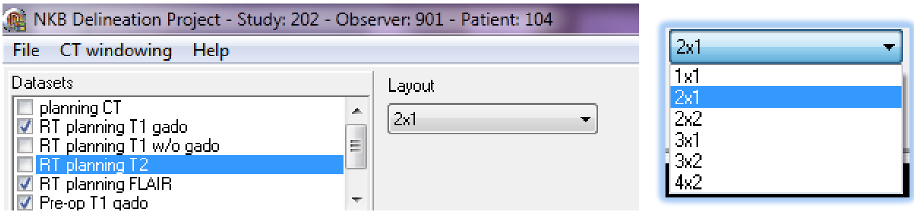
In both workflows, it was possible to return to modifying another ROI. However, in the traditional WF, it was discouraged since the CTV was automatically generated based on the GTV. As such, modifications to the GTV after CTV had been created did not update the CTV.

It is worth to mention, that even though the prototype was adapted for the contouring of CTV and edema, and the studies were conducted of these tasks, the results presented in the following chapters are only regarding the contouring of GTV, Surgical cavity, and CE residual tumor.

4.2.3.3 Data presentation

The layout of data display windows (part of the Graphical User Interface (GUI)) required modifications to accommodate the study needs. The original software was primarily developed for a single dataset based contouring (i.e., contouring directly on CT) while other modalities may have been available for reference. For our research, the layout was changed to accommodate eight datasets (Figure 4-4 bottom). The physician could define which datasets were to be visible, as well as the layout of them (e.g., 2 columns X 1 row, 3 columns X 2 rows) as seen in Figure 4-7.

In the initial software, contouring was enabled on the axial plane of the images, while the coronal and sagittal planes were presented at the right side (Figure 4-4 top). With introducing more datasets for contouring, it was not feasible to fit other planes in the primary GUI. Thus, the coronal and sagittal planes were visible as pop-up windows (Figure 4-8).



NKB Delineation Project - Study: 202 - Observer: 901 - Patient: 104

Datasets

☐ planning CT
☑ RT planning T1 gado
☐ RT planning T1 w/o gado
☐ RT planning T2
☑ RT planning FLAIR
☑ Pre-op T1 gado

Layout: 2x1

Figure 4-7 The dataset selection an layout modification section of the GUI

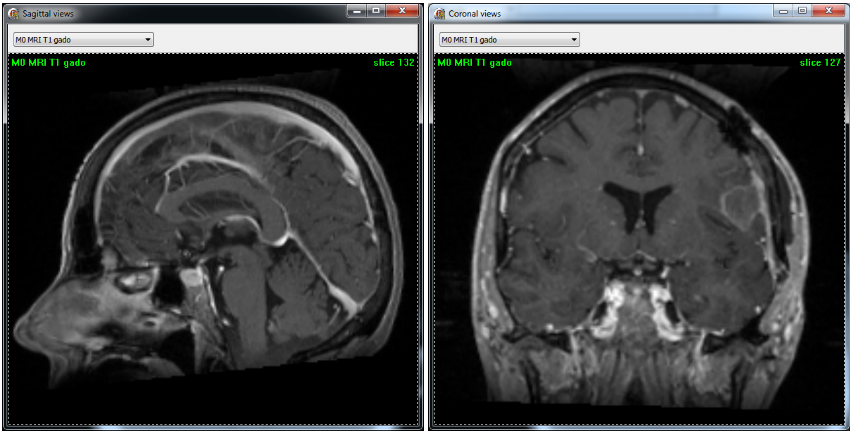


Figure 4-8 Examples of the sagittal and coronal pop-up windows

4.2.3.4 *Simplification of the GUI*

Another modification was to reduce the number of buttons on the GUI to limit visual distractions. Buttons that were duplicating mouse-based functionalities were removed, such as slice change, window-level change, zoom, and pan. The mouse-based direct manipulation of a dataset remained as it was in the original software (see Figure 4-9) - an image was divided into five regions, each of them enabling different function with right-mouse-click and hold-drag.

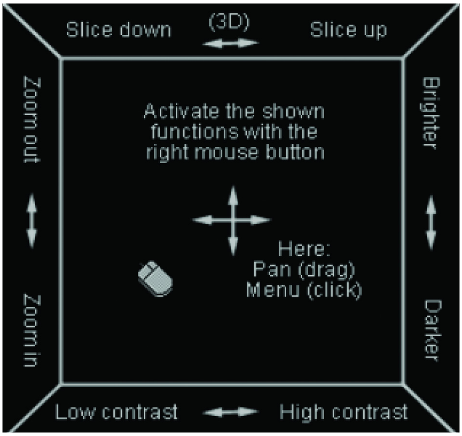


Figure 4-9 A guide to the mouse-based direct manipulation of an image dataset

4.2.3.5 Contouring tools

The number of contouring tools was kept to the minimum. All contouring tools were available for all ROIs. The contouring tools *create point* and *move point* were combined into one *polygon (point-based) drawing* tool. In addition, the *freehand drawing* and the *nudge* tools were made available.

Regarding automation tools, a *between slice contour interpolation* tool was also made available for users. Based on the assumption that the shape of the tumor is continuous, this tool is able to automatically generate a 2D contour based on the nearest inferior and superior contours (i.e., utilized contour information and not image information). No segmentation method based on the information of medical images, e.g., threshold based segmentation, was provided to physicians.

4.2.3.6 Integrated questionnaire

At the end of each contouring task, the physicians were presented with a questionnaire which consists of six standard questions of NASA Task Load index (NASA TLX) questionnaire (Hart and Staveland 1988) and two additional questions with commenting possibilities.

4.2.4 Prototype evaluations

Prior to conducting the studies, the prototype was evaluated with an experienced physician to identify the needs from the clinical perspective. The feedback was addressed in the tuning of the prototype, for instance, the CT and RT planning MRI T1-CE images were chosen as the default two image datasets according to the physician’s advices. In addition, usability testing pilots were conducted with a researcher experienced with medical imaging to identify possible usability flaws. The prototype was then iteratively improved based on identified problems.

4.2.5 Variations of the prototypes

Studies with the prototype were conducted subsequently in two locations: Germany and France, with one month in between. During the experiment in Germany it became clear that the polygon drawing is too slow, and consequently the contouring tools were adapted. Overview of the variation between contouring tools is shown in Table 4-4.

Table 4-4 Overview of segmentation methods

Study location	**WF type**	**Polygon drawing**	**Freehand drawing**	**Nudge**	**Within slice interpolation**	**Freehand deleting**	**Within slice deleting**
Germany	Traditional	Yes	no	no	no	yes	no
	Sub-region	Yes	no	no	no	yes	no
France	Traditional	No	yes	yes	yes	yes	yes
	Sub-region	No	yes	yes	yes	yes	yes

Differences on the GUI can be seen in Figure 4-10.

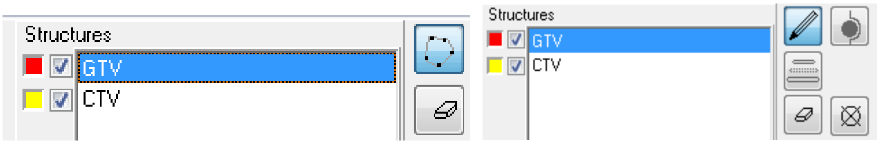


Figure 4-10 The differences of the contouring tools between two study locations of the traditional contouring workflow (left) study in Germany, (right) study in France

4.2.6 Interaction logging

The prototype involved key-level user interaction logging. The actions (mouse or keyboard) were recorded as time-stamped entries. The data was saved in a custom format that was later converted into text file (tab-separated values). The log file consisted on 21 columns. All data was captured with numerical values which had a corresponding label defined in the software code. The meaning of a column value depended on the type of user interaction. All modifications to the prototype required extending the user interaction logging.

Space limitations prevent listing the complete list of the log files. The log itself consisted of numerical values only. The meaning of those numbers depended on the type of event. Here is an overview of the data elements of different type of entries (Table 4-5). Based on these log entries, two new events were extracted: scrolling (series of slice changes), and pause (period of inactivity).

Table 4-5 Overview of the logged data details

Type of event	**Logged meta-data that was used in analysis**
Change the level	timestamp; dataset_ID; event_ID; state; value
Change the window	timestamp; dataset_ID; event_ID; state; value
Zoom	timestamp; dataset_ID; event_ID; state; zoom_level
Pan	timestamp; dataset_ID; event_ID; state;
Select a ROI	timestamp; event_ID; ROI_ID
Slice change	timestamp; dataset_ID; event_ID; slice_no
Drawing (contour line point event)	timestamp; dataset_ID; event_ID; state; position of the point drawn (x,y,z); drawing_type
Switch drawing tool	timestamp; event_ID; drawing_type
Interpolate a contour	timestamp; dataset_ID; event_ID; slice_no
Change nudge radius	timestamp; dataset_ID; event_ID; size
Change active viewer	timestamp; dataset_ID; event_ID
Open coronal plane dialog	timestamp; event_ID
Open sagittal plane dialog	timestamp; event_ID
Change dataset visibility	timestamp; dataset_ID; event_ID; visible%
Save work in progress	timestamp; event_ID

Parsing the log files was done with JavaScript. There were two types of interactions: singular and durational. A singular event resulted from a single (short) input (e.g., mouse click on a button), while durational event involved series of inputs (e.g., mouse click-and-drag for drawing). In addition, the durational events can be either active event or passive event. Active events occur one after another as the person can do one thing at the time. At the same time, the passive event runs in parallel on the background (e.g., “slice is visible”).

Table 4-6 Categorization of the events extracted from the log

Type	Active events	Passive events
Singular event	Interpolate; Select ROI; Switch drawing tool; Change nudge radius Save; Change active viewer; Slice change; Open coronal plane; Open sagittal plane; Change dataset visibility;	-
Durational event	Level change; Window change; Zoom; Pan; Scrolling; Pause	Slice visible; Dataset visible; Dataset active; Structure active;

4.2.7 Summary

The software prototype used to conduct user research for the study of contouring tasks of GBM was described. The prototype was adapted from existing contouring research software, and the modifications done were elaborated in this section. In total, there were four variations of the prototype - two types of workflows combined with two sets of contouring tools. An object-oriented model was used to describe the functionalities of the prototype. Details regarding the variations of the prototype were given as well. In addition, the details of the software interaction logging were given.

In section 4.1 we presented a set of design challenges to tackle during the design of the software. The research prototypes presented in section 4.2, we aim to get detailed understanding of physicians’ needs for general usability, navigation, workflow, and flexibility of interaction.

Chapter 5

Designing for sensemaking

This chapter is based on:

Aselmaa A, van Herk M, Laprie A, Nestle U, Götz I, Wiedenmann N, Schimek-Jasch T, Picaud F, Syrykh C, Cagetti L V, Jolnerovski M, Song Y, Goossens RHM (2017) Using a contextualized sensemaking model for interaction design: A case study of tumor contouring. Journal of Biomedical Informatics 65:145–158. doi: 10.1016/j.jbi.2016.12.001

Aselmaa A, Song Y, Goossens RHM (2016) User Interaction Visualization for Design Synthesis. In: Isenberg T, Sadlo F (eds) Poster Proceedings of the Eurographics Conference on Visualization (EuroVis 2016). The Eurographics Association, p 3 doi: 10.2312/eurp.20161128

The research presented in this chapter is narrowed to the sensemaking of tumor. The focus of this chapter is to explore the role of software design during sensemaking as well as the requirements for the design.

First, the impact of workflow with the corresponding differences in the user interface is explored (Section 5.1). Then, a two-step approach for incorporating sensemaking theory into design process is presented (section 5.2). Based on the conducted studies, both sensemaking and design insights are inferred regarding the tumor contouring task.

In addition, the visualizations that were developed for the analysis are presented in Section 5.3.

5.1 Interface design aligned with the micro-steps of tumor contouring workflow – an exploration

Radiotherapy treatment requires identifying the precise shape and location of a tumor (i.e., GTV) and can be cognitively challenging. A software solution that is in line with the cognitive steps of the GTV contouring workflow might make a better support for the physicians' reasoning process. The aim of this study was to explore the impact of the software which is designed in line with the cognitive micro-steps of tumor contouring (i.e., sub-structure based contouring), regarding both the contouring process and the results. A comparative study with eight physicians was conducted in two hospitals. The task completion times, use of medical image datasets and the GTV contours obtained with two software prototypes were compared. There were no significant differences in terms of task completion times and level of variability among physicians. At the same time, sub-structures based contouring resulted in on average 16% smaller GTV contours. Furthermore, fewer datasets were used during sub-structures based workflow. Implications of these findings are discussed and finally, recommendations for software designing and further research were made.

5.1.1 Introduction

Radiotherapy is an effective treatment against cancer. In radiotherapy planning, the exact shape and location of the tumor needs to be identified for an optimal treatment. This is done by contouring the visible borders of the tumor on 2D medical images following specific guidelines. For instance, in the delineation process of Glioblastoma Multiforme (GBM), the most commonly followed guidelines were published by the European Organization for Research and Treatment of Cancer (EORTC)(Stupp et al. 2009) and the US/Canadian Radiation Therapy Oncology Group (RTOG)(Colman et al. 2006). In addition, institutes use their own in-house guidelines that are developed based on international guidelines (e.g.,(Ghose et al. 2010; Gebhardt et al. 2014)).

Previously, in field studies (Aselmaa et al. 2013b), we observed that only the volumes that were required in the protocol were contoured. However, each volume may consist of several different sub-structures, which were described textually in the protocols. For instance, a recent ESTRO-ACROP guideline for the GBM contouring specified that "In macroscopically resected tumors Gross Tumor Volume (GTV) delineation should be based on the *resection cavity* (if present) plus any *residual enhancing tumor* on contrast-enhanced T1 weighted MRI, without inclusion of *peri-tumoural edema*" (Niyazi et al. 2016). In current practice, physicians often contour the GTV directly, by cognitively interpreting and fusing the information regarding different relevant sub-structures, i.e., resection cavity and residual enhancing tumor.

Besides contouring GTV directly, another workflow, which is more in line with the cognitive steps of physicians, might be, to separately contour the relevant sub-structures, and then computationally to fuse these contours to form the GTV. Such a workflow is built based on the Decomposition Hypothesis, which states that the execution of a complex task can be decomposed into a set of information-processing components and that these components combine unchanged in different task condition (Anderson et al. 2011).

Based on the Decomposition Hypothesis, the aim of this study was to explore the impact of the software which is designed in line with the cognitive micro-steps of tumor contouring, i.e., identifying sub-structures, regarding both the task process as well as the final results.

5.1.2 Materials and methods

5.1.2.1 Patient data

The datasets of three patients with resected primary glioblastoma were selected randomly (part of SPECTRO-GLIO clinical trial (2013a), consent forms obtained). Eight imaging datasets of different modalities were used for the study for each patient. They include pre-surgery datasets: MRI T1-weighted pre-gadolinium injection (MRI T1), MRI T1-weighted post-gadolinium injection (MRI T1-gado), and MRI FLAIR; and pre-radiotherapy datasets (acquired ~1 month after surgery): CT, MRI T1, MRI T1-gado, MRI T2-weighted (MRI T2), and MRI FLAIR.

5.1.2.2 Study participants and set-up

The contouring study was conducted in two hospitals (referred to as H1 and H2) in France and Germany, in total eight physicians participated. All physicians were informed about the details of the study and signed the consent form.

In the traditional workflow, the physicians were asked to contour the GTV_{trad} as consisting of *CE residual tumor* and *surgical cavity*. In the sub-structure based workflow, the task was to manually contour *surgical cavity*, and *contrast enhancing (CE) residual tumor*. The GTV_{new} was later calculated as a union of the manually contoured *Surgical cavity* and *CE residual tumor* by a computational algorithm. The differences of the workflows are illustrated in Figure 5-1

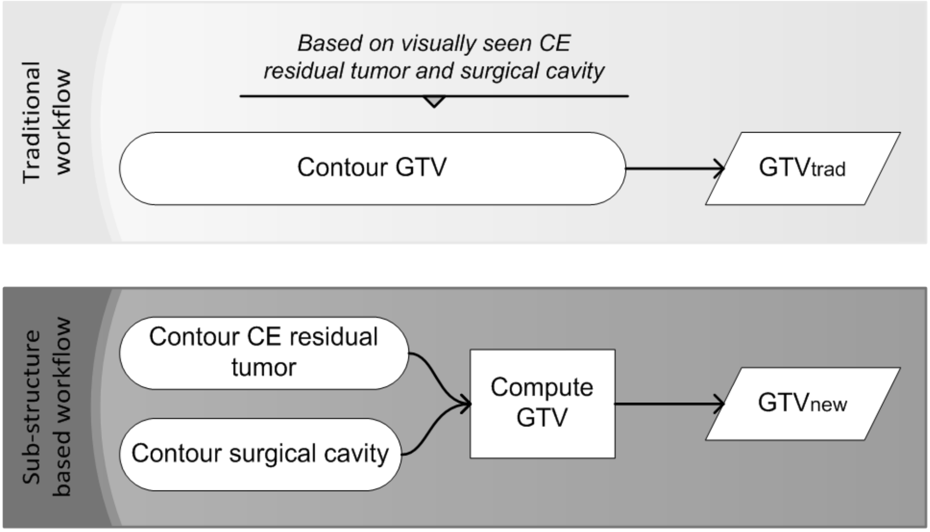


Figure 5-1 The two studied workflows of GTV contouring

5.1.2.3 Software prototype

To support the proposed research, a software prototype was developed based on the Big Brother software (Steenbakkers et al. 2005) which recorded user interactions into a log file. The software was extended to enable contouring on multiple imaging modalities. The datasets were presented in the following order from left to right on the graphical user interface: CT, MRI T1-gado, MRI T1, MRI T2, MRI FLAIR, pre-surgery MRI T1-gado, pre-surgery MRI T1, pre-surgery MRI FLAIR. In the beginning of the contouring task, the first two datasets (CT and MRI T1-gado) were visible for the physician.

The contouring tools and the graphical user interfaces of the original software were adapted. Besides basic contouring tools, the modified software allowed changing the layout and the number of visible imaging datasets at any given point. After conducting the study in the first hospital, the contouring tools were adapted to allow faster interactions (e.g., the point-based drawing tool was replaced with a freehand drawing tool).

Table 5-1 Overview of the differences in the drawing tools between hospitals

Hospital	Workflow type	Point-based drawing	Freehand drawing	Nudge	Within slice interpolation	Freehand deleting	Within slice deleting
H1	Traditional	Yes	No	no	no	yes	no
	Sub-structure	Yes	No	no	no	yes	no
H2	Traditional	No	Yes	yes	yes	yes	yes
	Sub-structure	No	Yes	yes	yes	yes	yes

In the beginning of the study, physicians were introduced to the software and were given time to try out different functionalities of the software on a test case. Once they were familiar with the software, they started with the given contouring tasks. The sequence of cases was different among physicians but same for both approaches per physician.

The software incorporated automatic dataset selection and immediate display for the *CE tumor* (two datasets) and *Surgical Cavity* (three datasets) contouring tasks. For the *CE tu*mor, pre-radiotherapy MRI T1-gado and MRI T1 were presented side by side. For the *Surgical cavity* the pre-radiotherapy CT was displayed in addition to the two MRI T1 datasets.

5.1.2.4 Contouring time

The total completion time of a task was calculated based on the logged data for each case finished by a physician. Then a mean case completion time was calculated for each case regarding two hospitals.

For each image dataset, its activeness and visibleness durations were summarized as a percentage of the overall case completion time. Then, the mean activeness and visibleness durations of the datasets were calculated per structure and observer. Activeness of a dataset was defined as the period of time from the first interaction with the dataset until an interaction with another dataset or with the graphical user interface elements occurred. Visibleness of a dataset was defined as a period of time of it being visible on the screen.

5.1.2.5 Contour analysis

All structures from both approaches were analyzed separately based on the methods described by Steenbakkers et al. (2005) First, the volumes of all the structures were calculated. For all the structures, the 50% agreement surface ($V_{50\%}$) was calculated. The $V_{50\%}$ represents the enclosing volume that at least 50% of physicians included in the structure. For each point describing the $V_{50\%}$ surface (about 280 points/cm^2), the perpendicular distance was measured. The overall variation in distance to all points describing the $V_{50\%}$ surface of a structure was expressed in an overall standard deviation (overall SD).

5.1.2.6 Statistical analysis

Paired samples t-test was used to examine the impact of the sub-structure based workflow on the task completion durations within the two hospitals, as well as on GTV volumes. All statistical tests were performed using the SPSS for windows software (version 22). Due to the exploratory nature of the work, the $p < 0.1$ was deemed significant.

5.1.3 Results

The calculated durations and volume measures are displayed in Table 5-2.

Table 5-2 Calculated measures. H1 = Hospital 1; H2 = Hospital 2; VMean = Mean volume; V50% = 50% agreement volume; Overall SD = average standard deviation of the observer surface to the V50% surface distance; CI = conformity index; CVvol = coefficient of variation of the volumes.

	C-1		**C-2**		**C-3**	
	GTV_{trad}	**GTV_{new}**	**GTV_{trad}**	**GTV_{new}**	**GTV_{trad}**	**GTV_{new}**
Mean $duration_{H1}$	9min 48s	11min 29s	21min 33s	25min 22s	15min 14s	19min 25s
Mean $duration_{H2}$	6min 21s	7min 58s	11min 3s	11min 55s	10min 13s	10min 43s
V_{Mean} (cc)	14.0	12.3	29.8	23.1	23.2	19.4
$V_{50\%}$ (cc)	13.6	12.4	29.8	23.8	24.0	20.0
Overall SD (mm)	1.6	1.6	1.8	1.7	1.3	1.2
Mean CI	0.82	0.84	0.83	0.83	0.89	0.87

	C-1		**C-2**		**C-3**	
	GTV_{trad}	**GTV_{new}**	**GTV_{trad}**	**GTV_{new}**	**GTV_{trad}**	**GTV_{new}**
CV_{vol}	14.1%	15.3%	11.1%	12.1%	5.7%	9.2%

5.1.3.1 Contouring times

The contouring times in hospital H2 were less in all cases and in both workflows. The differences in contouring times between the two hospitals can be largely accounted for the change of contouring tools. The mean duration per case and hospital showed a tendency towards an increase (see Table 5-2).

The task completion of a case took longer for the sub-structure based contouring on average by 1min 37s, 1min 59s and 1min 53s, but did not reach significance (p = 0.170, p = 0.296, p = 0.339) respectively for the cases C-1, C-2, and C-3. The contouring time of individual physicians between the two settings decreased in nine occasions on average by 2min 22s and increased on 15 occasions on average by 4min 49s (Figure 5-2).

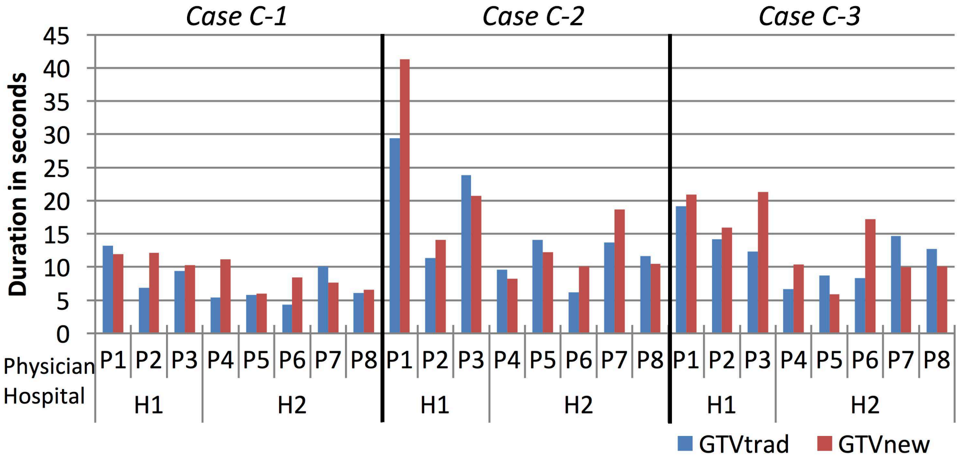


Figure 5-2 The task completion times of all the physicians. Physician P1-P3 were from hospital 1 (H1), P4-P8 from hospital 2 (H2).

5.1.3.2 Contoured volumes and variation

The mean volumes of the GTV from both approaches, as well as the mean CV of volumes, and the overall SDs are presented in Table 5-2 regarding each case, respectively. Distribution of the volume sizes is shown in Figure 5-4.

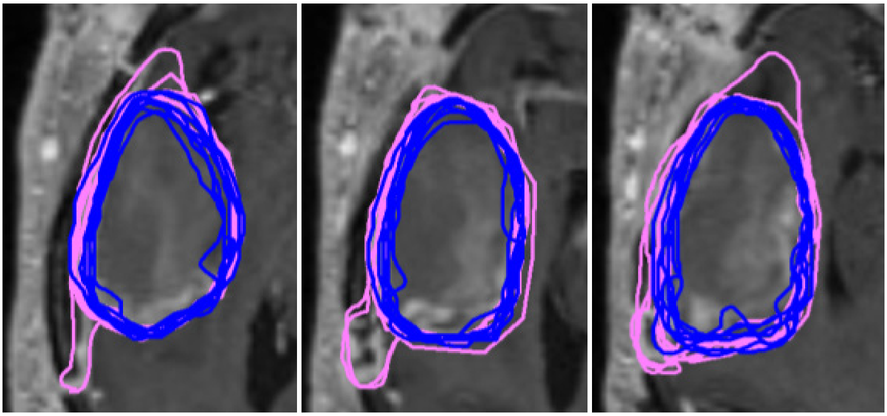


Figure 5-3 Examples of the resulting contours from eight physicians of the case C-2 on three axial planes overlaid on MRI T1-gado; pink = GTV_{trad}, blue = GTV_{new}

The individual GTV_{new} volumes were consistently smaller compared to GTV_{trad} for the three cases, on average by 1.7cm^3 (11%, p=0.057), 6.7 cm3 (22%, p<0.0005), and 3.9 cm3 (16%, p=0.002) respectively. The calculated 50%-agreement volume of GTV_{new} of each case were smaller by 1.2cm^3, 6.0cm^3, and 4.0cm^3 respectively. At the same time, the CV_{vol} of the GTV_{new} increased by 1.2%, 1.0%, and 3.5%. The CI and the overall SD showed no significant difference between the GTV_{new} and the GTV_{trad}.

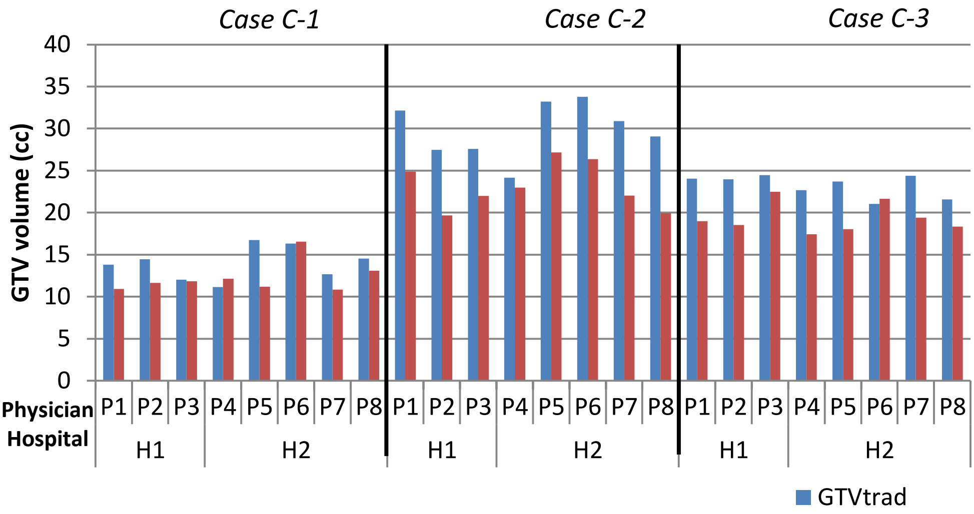


Figure 5-4 The individually contoured GTV volumes by the eight physicians (P1-P8).

5.1.3.3 The use of multimodal images

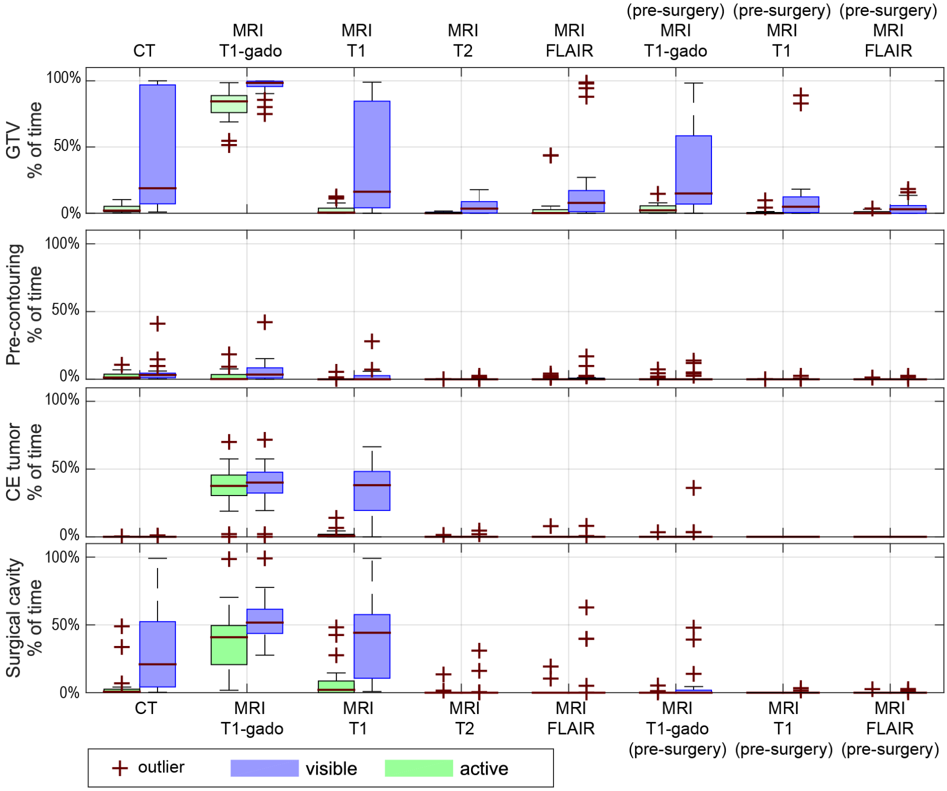


Figure 5-5. Activeness and visibleness times of image datasets during the contouring in traditional (GTV, top graph) and sub-structure based workflow (Pre-contouring, CE tumor, and Surgical cavity). The imaging modalities are in the same order (from left to right) as they were displayed within the software prototype. Outlier points are identified as below or above the ± 1.5 of interquartile range, corresponding to approximately ± 2.7σ and 99.3 percent coverage if the data.

In Figure 5-5 the activeness and visibleness times of all image datasets are presented regarding each contoured structure. Since in the sub-structure based workflow the physician was allowed to start by contouring either one of the two sub-structures (CE tumor or Surgical cavity), the pre-contouring time is indicated separately.

Activeness of different imaging datasets showed a clear association with the structures being contoured. During contouring of $\mathrm{GTV}_{\mathrm{trad}}$ more image datasets were viewed than during the contouring of structures for the sub-structure based combined. The activeness remained similar, the radiotherapy planning MRI T1-gado being used primarily, as expected.

5.1.4 Discussion

A study comparing the traditional workflow against the sub-structure based workflow of GTV contouring was presented. This approach intended a shift the focus from “identify abnormalities” to “identify [a specific sub-structure]”, which in turn can be further supported by detailed guidelines or interface design elements (e.g., computer-interpretable clinical guidelines (Peleg 2013)). Such an approach is expected to reduce the need of cognitively fuse information, while ensuring that the main aspects of a tumor are being explicitly considered.

5.1.4.1 Influences on the contouring process

The design of the interface in the sub-structure based workflow proposed the datasets and layout for the contouring of each of the sub-structure. Even though it was possible to change the pre-set layout and datasets, only few physicians chose to do so. For instance, during the contouring of the *CE tumor* only in two cases out of the 24, the pre-radiotherapy CT was additionally viewed (visible for longer than 1 second). During the contouring of GTV_{trad}, often datasets other than the ones used for contouring were consulted. This can be seen in Figure 5-5 as the difference between the duration of a dataset being visible compared to being active. In many cases the CT was visible for the whole GTV_{trad} contouring process. As the CT was the first dataset presented on the GUI on the start of the task, it may have influenced the physicians to keep it visible. These changes in the interactions with the medical images indicate that the interface design impacts physicians’ behavior.

The least used radiotherapy planning dataset was MRI T2. This can be explained with the availability of MRI FLAIR, which presents similar information as a MRI T2 but suppresses the signal of cerebrospinal fluid thus allowing better demarcation of edema from brain tissue (Essig et al. 1998).

The visibleness of imaging datasets during the contouring of the sub-structures was typically limited to two or three datasets. At the same time, during the GTV_{trad} contouring, larger variability was observed in terms of which datasets were visible. The sub-structure based contouring focused the physicians towards the task of sub-structure contouring and reduced the need to seek for further information and fusing information from more modalities.

Previous research indicated that breaking a task into micro-tasks results in longer overall task completion times, but higher quality outcomes (Cheng et al. 2015). Furthermore, Cheng et al. stated that "micro-tasks can help people complete high quality work in interruption-driven environments." Interruptions are common during radiotherapy treatment planning (Aselmaa et al. 2013a), thus the presented sub-structure based approached could allow decreasing the effect of interruptions. However, further investigations are required.

5.1.4.2 Influences on the contouring results

As expected, there were variations among the manually contoured GTV_{trad}. One of the reasons for the variation is the physicians' subjective preferences (e.g., some tend to contour larger volumes than others). Regarding the GTV_{trad}, the extent of the surgical entry wound inclusion had an impact. Expansion of malignant glioma to the bony structures as well as to the soft tissues outside cranium occurs rarely and this structure thus can be considered a natural barrier (Bokstein et al. 2008). In few cases, the borders of the GTV slightly protruded over the boney region. As such, incorporating computer-aided solution to allow adapting GTV to the bony structures can be a requirement for all radiotherapy treatment planning systems.

In one of the three cases, a region included within the GTV_{trad} by some of the physicians was the contrast enhancing dura. However, the contrast uptake is more likely to represent reactive changes of the dura than tumoral invasion (Wilms et al. 1991). At the same time, the sub-structure based approach resulted in GTV_{new} contours that excluded such region.

Large variations were observed regarding the volumes and shapes in the manually contoured S*urgical cavity* (mean CI = 0.78, mean overall SD = 1.9 mm), and *CE residual tumor* (mean CI = 0.49, mean overall SD = 3.9 mm). One of the reasons for those variations was the inclusion or exclusion of either of these structures within each other. In the case of *Surgical cavity* and *CE residual tumor*, it was cognitively difficult for the physicians to separate one from the other since in the current clinical practice, there is no need to do so. Thus, there was limited or no previous experience in such contouring tasks. Therefore, having more detailed contouring instructions and also more thorough training for contouring these structures could improve the consistency and potentially would lead to some reduction of variations among physicians.

Within all three cases, GTV_{new} was slightly smaller than GTV_{trad} ($p=0.057$, $p<0.0005$, and $p=0.002$ respectively). Physician following the sub-structure based workflow on average reached 16% smaller volume (4 cm3) than contouring the GTV directly. This difference could be led by the effect of the mental fusion process in the traditional workflow. However, it should be taken into consideration that reduction of volumes may introduce additional risk of not including all tumorous tissues. As there is a high level of uncertainty of where the tumor really is, physicians generally prefer to include all areas of suspicions.

5.1.4.3 Design implications

In the contouring process, the contouring guidelines, which specify the relevant regions of interest and the procedures for defining target volumes, are used. As we stated before, in GBM contouring, the most commonly followed guidelines were published by the EORTC (Stupp et al. 2009), the RTOG (Colman et al. 2006) and more recently ESTRO-ACROP committees (Niyazi et al. 2016). In addition, institutes use their own in-house guidelines that are developed based on international guidelines (e.g., (Ghose et al. 2010; Creak et al. 2011)). For instance, according to a Canadian survey more than half of responding physicians indicated that they follow in-house guidelines for GBM treatment (Ghose et al. 2010). While there seems to be benefit of aligning the software to the detailed workflow, it is not feasible to design for all protocols. As such, a more generic solution is required that has minimum user involvement in terms of administration and customization.

Considering that there are many variations in contouring protocols within and between hospitals, there are implications to the design. First, the user interface has to enable choosing the relevant protocol and prepare the “working space” accordingly, for example by providing a list of relevant volumes to be contoured. Second, with the increasing complexity of treatment plans and contouring tasks, there is a need for basic Boolean operations for creating and modifying volumes. Third, the clinical knowledge need to be increasingly integrated into the target volume creation. For instance, many software solutions have the option to automatically exclude the bone region from a volume. In our view, many more such clinical recommendations should be incorporated to the software. For instance, in this presented study, with the sub-structure based approach the contrast enhancing dura, which should be excluded from the GTV, was more likely excluded compared to the classical approach.

The study presented here was experimental in its nature. The sub-structure based approach assumed that it is sufficient to identify the sub-structures of GTV. However, during post-hoc discussion with physicians it was pointed out that in clinical practice, possibility for reviewing or correcting the calculated GTV is usually necessary. This shortcoming of the study should be addressed in future studies.

5.1.5 Conclusion

The influence of the interface design aligned to the traditional workflow or to the sub-structure based workflow of GBM GTV contouring were evaluated. In the sub-structure based workflow, GTV was calculated based on the manually contoured sub-structures: *surgical cavity*, and *contrast enhancing (CE) residual tumor.*

Compared to the traditional workflow of contouring GTV directly, the sub-structure based workflow results in slightly smaller volumes (decrease by 11%, 22%, 16% for the three cases studies). There were no significant differences for the task completion time in relation to the type of workflow. However, the image datasets were significantly less interacted with in the sub-structure based approach decreasing the need to cognitive information fusion.

Our findings suggest that an approach where physicians contour the underlying sub-structures of GTV rather than directly contouring GTV leads to similar level of variation. However, further research is needed regarding the influence on the efficiency as well as to identify medical relevance of the observed differences.

5.2 Using a contextualized sensemaking model for interaction design

Sensemaking theories help designers understand the cognitive processes of a user when he/she performs a complicated task. This section introduces a two-step approach of incorporating sensemaking support within the design of health information systems by: (1) modeling the sensemaking process of physicians while performing a task, and (2) identifying software interaction design requirements that support sensemaking based on this model. The two-step approach is presented based on a case study of the tumor contouring clinical task for radiotherapy planning. In the first step of the approach, a contextualized sensemaking model was developed to describe the sensemaking process based on the goal, the workflow and the context of the task. In the second step, based on a research software prototype, an experiment was conducted where three contouring tasks were performed by eight physicians respectively. Four types of navigation interactions and five types of interaction sequence patterns were identified by analyzing the gathered interaction log data from those twenty-four cases. Further in-depth study on each of the navigation interactions and interaction sequence patterns in relation to the contextualized sensemaking model revealed five main areas for design improvements to increase sensemaking support. Outcomes of the case study indicate that the proposed two-step approach was beneficial for gaining a deeper understanding of the sensemaking process during the task, as well as for identifying design requirements for better sensemaking support.

5.2.1 Introduction

Health information systems (HIS) refer to computer based information systems (i.e., software and hardware) used in healthcare settings (Yusof et al. 2008). HIS were initially developed for patient care and administrative purposes, but are now being gradually extended to different areas of healthcare planning (Haux 2006). With the continuously growing amount of digital data, treatment planning relies more and more on software solutions. At the same time, the effectiveness and efficiency of those software solutions depend on whether they can successfully combine the physicians' expertise with the computing power, and whether they fit well into the clinical workflow. Among the ongoing research activities for improving HIS, there is an increased interest in supporting physicians' cognition while they are performing clinical tasks. This indicates the growing role and the importance of cognitive science within HIS design (Patel and Kannampallil 2015). However, many of current solutions only offer limited support to typical cognitive tasks in the clinical domain, such as decision making and prevention of medical errors (Viitanen et al. 2011).

5.2.1.1 Background

Sensemaking is the process of creating an understanding of a concept, knowledge, situation, problem or work task, often to inform an action. It is a prerequisite for problem solving and decision making (Zhang and Soergel 2014) as such: *"better understanding of human sensemaking processes is critical for understanding how information processed through information systems is appropriated by human users and converted into knowledge and resulting action and performance"* (Malhotra 2001). In general, sensemaking can be seen as the process of searching for a frame (also referred to as knowledge, a mental model, a representation, or a structure) and encoding data into that frame to answer task specific questions (Russell et al. 1993). Throughout a task, one is *"facing gaps, building bridges across those gaps, evaluating outcomes and moving on"* (Dervin 1998). Furthermore, the interplay between frames and data is bidirectional as *"frames shape and define the relevant data, and data mandate that frames change in nontrivial ways"* (Klein et al. 2006b).

Most sensemaking models consist of loops or cycles, which indicates that sensemaking is generally seen as an *iterative* process. This process usually starts from a *goal*, and takes place through the use of *data*, to build and update the *frames* iteratively until one has reached a satisfactory outcome. Furthermore, *gaps* (i.e., discrepancies between *data* and *frame*, or between *frames*) are typically seen as the triggers behind the sensemaking activities. The driving force for the sensemaking activities is to explain the *gaps*, resulting in updating the *frames* or *data*. As such, in a broad understanding, sensemaking connects the *data* and *frame* through a series of sensemaking activities (i.e., sensemaking loops) to build and update the *frame* according to a specific task goal as illustrated in Figure 5-6.

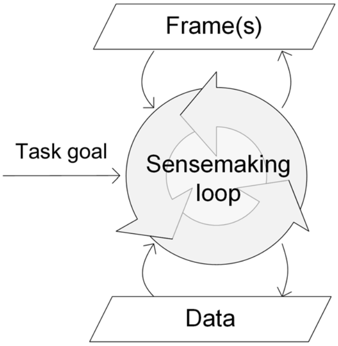


Figure 5-6. A generalized sensemaking model. The frame represents a cognitive structure of a concept, knowledge, etc. Data is being iteratively fitted to the frame through the sensemaking until the task goal is achieved to a satisfactory level.

Sensemaking theories have been developed for having a better understanding of the cognitive process mainly in four fields (Dervin and Naumer 2009): Human-Computer Interaction (Russell et al. 1993); Cognitive Systems Engineering (Klein et al. 2006b, 2007); Organizational Communication (Weick 1995a) and Library and Information Science (Dervin 1998). In the past decade, research activities regarding understanding sensemaking process and applying sensemaking theory in different fields has been increasing. For instance, Russel et al. held two workshops on sensemaking at two consecutive Conference on Human Factors in Computing Systems (CHI 2008 (Russell et al. 2008) and CHI 2009 (Russell et al. 2009)). Such an increase of interest can be accredited to multiple factors: the explosion of information in the Web; the increased number of projects in library and information sciences; the needs to help people make sense of the multitude information resources available and in response to the growing interests from various funding agencies in improving homeland security, emergency response, and intelligence analysis (Pirolli and Russell 2011).

The concept of information foraging, consisting of information seeking, gathering, and consumption (Pirolli and Card 1999), is closely associated to sensemaking. For instance, Pirolli and Card (Pirolli and Card 2005) developed a notional sensemaking model that described intelligence analysis process. This model consisted of both foraging loops and sensemaking loops. Depending on the sensemaking theory, information seeking can be seen as a part of or strongly coupled to sensemaking. As such, research on information seeking behavior can bring relevant insights for comprehending sensemaking. For instance, Kannampallil et al. (2013) observed that the information seeking process was exploratory and iterative, and it was driven by the maximized information gain from information sources. Such a view of information seeking is very similar to sensemaking, which can be seen as an iterative information processing task, during which one attempts to reduce the cost of operations (Russell et al. 1993).

In the research area of applying sensemaking in the healthcare context, Mamykina et al. (Mamykina et al. 2015b) developed a theoretical sensemaking framework in a study of chronic disease (diabetes) management. Such a sensemaking based framework can be used as a new analytical lens that could enrich the existing scholarship and suggest new directions for research and for the design of technological interventions. Sensemaking approaches can also be beneficial in shaping and framing research about HIS (Bossen and Jensen 2014). Besides, collaborative sensemaking had been applied in hospital emergency department setting (Paul and Reddy 2010), nursing (Kristiansen et al. 2015), and online health forums (Mamykina et al. 2015a). Other specific areas of collaborative sensemaking that have been investigated are: team collaboration (Reddy and Jansen 2008; Leykum et al. 2015), handoffs (Sharma 2009), etc.

Although there is a range of sensemaking models available in different domains and contexts, most of them focus on describing and explaining the sensemaking process. Literature review indicates that few studies systematically used sensemaking models to identify requirements for HIS, or more specifically to describe how to support the design of software for HIS from the sensemaking perspective. In many cases, HIS designers have to use their intuition and experience to interpret and apply the theoretical sensemaking in the HIS software design, thus it is difficult to keep a holistic view of sensemaking process of a given task as well as to extract detailed design requirements from sensemaking for each step of the task.

5.2.1.2 Research approach

The aim of this section is to introduce an approach that uses a (contextualized) sensemaking model to support interaction design of HIS software. Using a case study of tumor contouring task for radiotherapy treatment planning, we formulate the proposed approach in two steps (Figure 5-7): (1) using sensemaking theory and contextual knowledge to develop a contextualized sensemaking (C-SM) model. This model gives designers a holistic view of sensemaking process as well as a deeper understanding of different moments that sensemaking takes place while the user uses a software solution for a given task; (2) analyzing the software interactions (patterns) using this C-SM model in order to generate detailed insights of the sensemaking process and to identify requirements for the design.

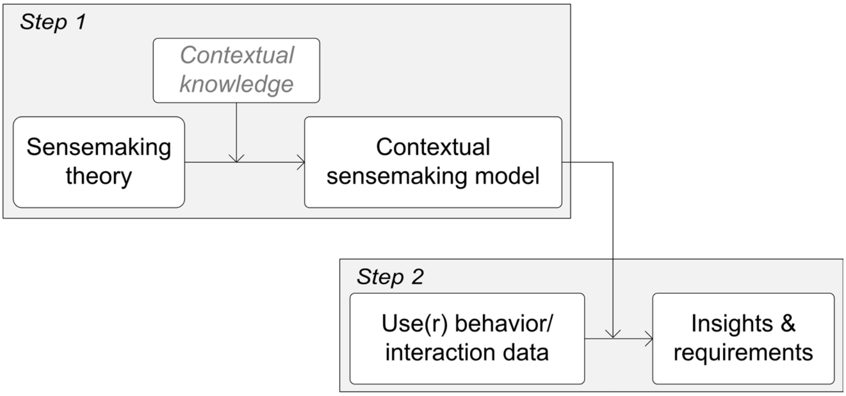


Figure 5-7. The proposed two-step approach.

5.2.2 Modeling sensemaking in the context

In this section, based on the previously described generalized sensemaking model, we develop the C-SM model by incorporating contextual knowledge regarding the task, its clinical context, and the software interactions which are crucial for completing the task. The aim of the C-SM model is to identify the relations between the task process and the interactions with the software throughout the sensemaking process.

5.2.2.1 The task – tumor contouring for radiotherapy planning

Radiotherapy is a medical treatment against cancer, during which a high dose of radiation is delivered to the tumor while attempting to spare the normal tissue. Since tumors are within the human body, medical images (e.g., Computed Tomography (CT) scans or Magnetic Resonance Imaging (MRI) scans) are usually the primary data source for the treatment planning. These images, which represent (part of) the three-dimensional (3D) human body, are presented on the computer screen as a set of 2D images (i.e., slices). In radiotherapy treatment planning, physicians navigate through these 2D images to construct the mental 3D model of the anatomy (Varga et al. 2013) for different tasks.

Radiotherapy treatment planning has a complex interdisciplinary workflow that involves multiple clinicians (e.g., radiologists, radiation oncologists, medical physicists) and a series of tasks (e.g., medical image acquisition, radiation dose plan validation). This procedure usually takes several days, and often multiple software solutions (Aselmaa et al. 2013b) are used. Once a patient has been diagnosed with cancer and radiotherapy has been advised as (part of) the treatment, multiple modalities of medical images are acquired (e.g., CT, MRI, etc.). Each imaging modality provides unique clinical information relevant for the treatment planning. Images from different modalities are then co-registered in the same coordinate space to allow easier extrapolation of information at the same location. This is followed by one of the critical tasks that significantly influences the outcomes of the treatment - identifying the location and the shape of the tumor (i.e., the contouring task). This is achieved by drawing 2D contours on each relevant slice. A set of these drawn 2D contours represents a 3D *volume* of a certain aspect of the tumor. In the radiotherapy planning, different types of volumes are needed and one of the important volumes, Gross Tumor Volume (GTV), represents the macroscopic spread of the tumor (i.e., what can be seen as tumorous tissues with naked eye) (Burnet 2004). Other volumes are then identified based on the GTV by incorporating medical knowledge regarding the expected tumor spread (i.e., the non-visible tumor), and uncertainties of the treatment delivery (e.g., possible movements of the patient). Once all the relevant volumes are contoured and validated, physicians may start radiation dose planning and validation.

The advancements of technology in the past decades have made it possible to deliver the radiation to very complex shapes (Nutting et al. 2000). Therefore, accurately identifying all the relevant volumes is critical for an optimal treatment. However, tumor contouring is considered to be the weakest link in radiotherapy planning (Njeh 2008), and large interobserver variabilities among physicians have been identified in several case studies (e.g., Figure 5-8). For example, in a study of contouring the GTV of a Glioblastoma Multiforme (GBM, a very aggressive type of primary brain tumor), the average relative standard deviation (standard deviation over the mean) of the Dice-Jacaard coefficient of the GTV varied from 0.39 to 0.64 for nine cases (Wee et al. 2016). This indicated a high interobserver variability among physicians, thus the final treatment plan highly depends on the judgement of individual physicians.

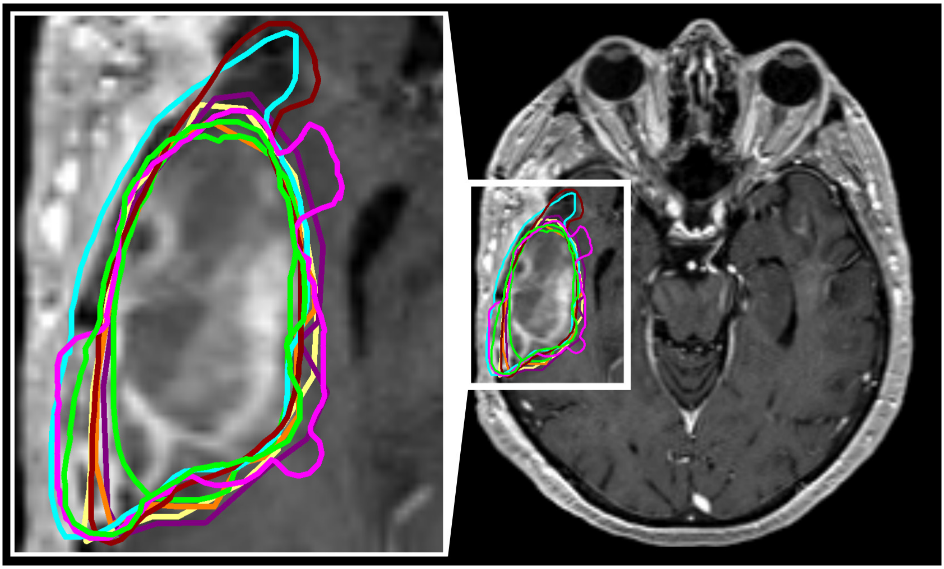


Figure 5-8. Example contours of a GTV on one 2D slice as contoured by eight physicians (each in different color) in a case of GBM, overlaid on MRI T1-weighted image with contrast enhancement. High interobserver variability can be observed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The contouring task is cognitively demanding as there are multiple variables that the physicians need to take into consideration (Aselmaa et al. 2014). The main challenge of the physicians is to distinguish between the tumorous tissues and the normal tissues. The boundaries of the tumor on the medical images are often not clear, thus the physician needs to obtain and synthesize additional data in combination with their knowledge and experience in order to reach a decision. The additional data can be either from the neighboring 2D images, or from other medical image datasets in different modalities. Besides, the treatment details (e.g., palliative or curative treatment, influence of chemotherapy), and tumor characteristics (e.g., proximity to organs at risk, level of infiltration) may influence the reasoning as well. In this cognitively demanding process, sensemaking can be seen as the underlying process that the physicians are engaged while contouring the GTV, and through which they try to overcome the complexity and uncertainties in order to complete the task. As such, having a better understanding of the sensemaking process could enable reaching a better design of the software solution used for the contouring task.

5.2.2.2 Phases of the task

Task phases (stages) have an impact on the types of sources used, judgement of relevance and information search strategies (Kim and Soergel 2006). To acquire the contextual knowledge, besides literature studies, we conducted observational research studies at Department for Radiation Oncology, University Medical Center Freiburg, Germany and Département de Radiothérapie, Institut Claudius-Regaud, Institut Universitaire du Cancer de Toulouse-Oncopole, France. In the one-week long research, we interviewed five physicians and observed more than five tumor contouring tasks that were completed using different software solutions. Such observational research helped the study in: (1) understanding the workflow and the relations among different tasks in the workflow; (2) familiarizing with the context of the GTV contouring task and (3) generating a qualitative description of the task.

Through the observational research study and workflow analysis (Aselmaa et al. 2013b), three main task phases – named the familiarization phase, the action phase and the evaluation phase – were identified in the GTV contouring process as shown in Figure 5-9. In the familiarization phase, the physician becomes familiar with the task and the data, and identifies the gaps between data and frames. During the action phase the physician is engaged in the interactions that directly contribute to the task completion (e.g., contouring, navigating). In the evaluation phase, the physician evaluates the outcomes (i.e., contours) against the information perceived from the medical images and his/her medical knowledge. The gap identification during this phase can be either hypothesis based (based on knowledge) or data based (based on what is seen). When a gap is identified, the physician returns to the action phase to make the necessary corrections.

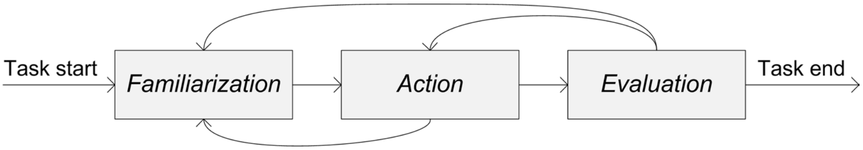


Figure 5-9. Three task phases in the GTV contouring.

The boundaries among different task phases are fuzzy and the sequence of them is not always linear. The familiarization phase occurs mostly at the beginning of the task. Additional rapid familiarizations may take place when the physician performs the action or evaluates results (e.g., data modification or presentation change). However, this type of familiarization is more related to visual perception than to specific software interactions. The action phase can be determined based on the interactions which are performed to directly support the goal of the task. The evaluation phase is often intertwined with the action phase. For instance, in the evaluation, when the physician identifies a discrepancy between the contour and the image, he/she usually corrects the contour immediately (i.e., perform actions) and then continues with evaluation.

5.2.2.3 The C-SM model of the task

In order to develop a sensemaking model suitable for describing the context of tumor contouring, the generalized sensemaking model described in Figure 5-6 was extended and adapted to the software's use context first. Here, the model developed by Zhang and Soergel (2014) was partly adopted as it describes individual sensemaking while incorporating ideas from learning and cognition. In their model, they identified seven key sensemaking activities: task analysis, identification of gaps (data or frame), information (data or frame) seeking (exploratory or focused), building frames, fitting data into frames, updating frames, and preparing task output as illustrated in Figure 5-10.

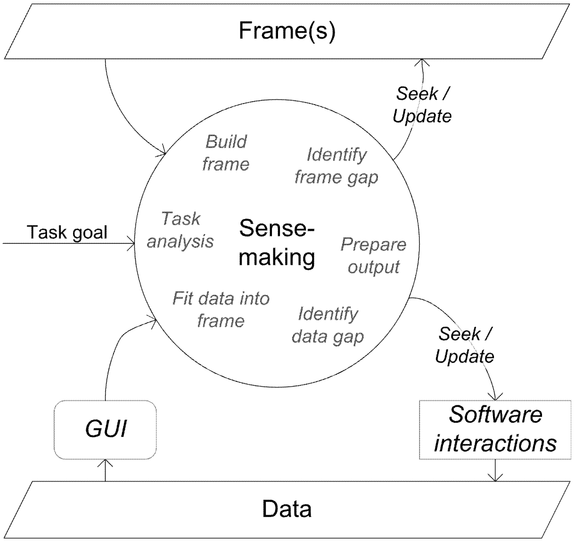


Figure 5-10. Generalized sensemaking model in the context of software use, including the sensemaking activities as identified by Zhang and Soergel (2014).

Identification of gaps (*data gap* or *frame gap*) is seen as the central activity of sensemaking. After the gap is identified, *information seeking activities* take place to find a *data* or *frame* that bridges the *gap*. The *gap bridging* *activities* take place through *building frame* and *fitting data into frame* in symbiosis. Throughout this process, one is *updating frame (i.e., knowledge)* and *preparing task output*. The task output is generated by *updating the data*. When the sensemaking is taking place through the use of software, the information seeking from data and generating task output is achieved through software interactions. At the same time, all the data is presented on the Graphical User Interface (GUI) and perceived based on this presentation.

Based on the identified GTV contouring task phases (as illustrated in Figure 5-9) and the generalized sensemaking model in the context of software use (as shown in Figure 5-10), the C-SM model could be generated. First, the types of context specific frames were identified. Then, the primary connection points with the software solution (i.e., the GUI and the types of software interactions) were identified and positioned within the three task phases that were described previously. The resulting C-SM model is illustrated in Figure 5-11.

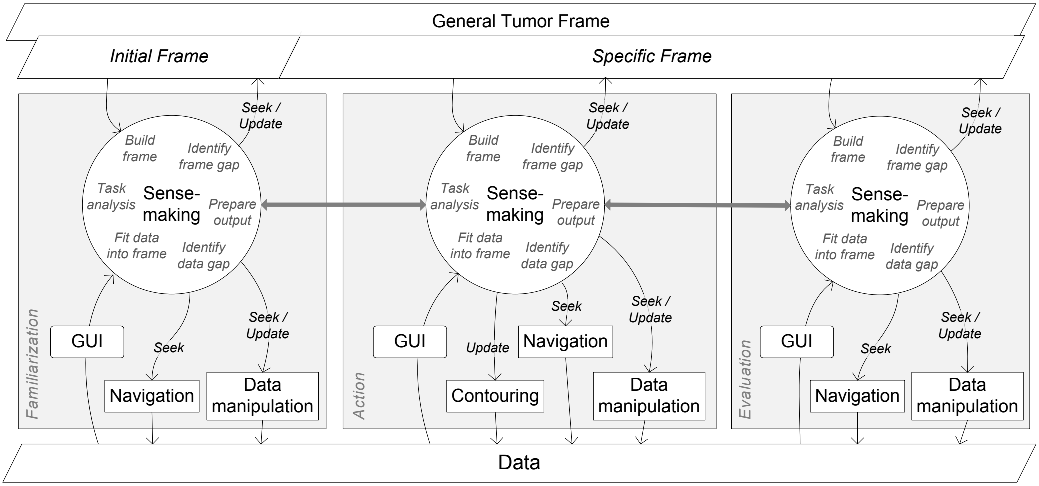


Figure 5-11. The contextualized sensemaking model (C-SM) of the tumor contouring task.

During tumor contouring, the frames involved represent primarily instances of a *general tumor frame* as the parallelogram at the top of Figure 5-11. The *general tumor frame* represents the physician’s knowledge, clinical experience, expectations of the tumor and it is iteratively updated throughout each sensemaking iteration. For each case, an *initial frame* is generated based on the data of the case. Throughout the task, this *initial frame* gradually evolves towards a *specific frame* through a series of sensemaking activities.

The sensemaking process results in seeking for a frame, updating the frame, or in an intent for performing interactions with the software solution. The interaction with the software is achieved by using a mouse, a keyboard, etc. Once the input is given to the software, the results can be perceived through the GUI. The primary software interactions during the contouring task are for navigation, data manipulation and contouring. Through these interactions the data or its presentation is changed, allowing the physician to view and evaluate the outcome on the GUI for continuing with the sensemaking process. The primary output of the GTV contouring task is the contour (stored digitally as data) that represents the *specific frame* in an externalized form.

5.2.3 The case study

In order to gain a deep understanding of the sensemaking process and to get detailed information about the software interactions involved in the process, a case study of GTV contouring of the GBM tumor was conducted. The GTV contouring task was chosen for the study for two reasons: (1) the GTV is used as a basis for generating other volumes in radiotherapy treatment planning and (2) the task is cognitively challenging by nature. This section describes the setup of the case study, the materials and methods used in the study, and the detailed overview of the software interactions..

5.2.3.1 Materials and methods

5.2.3.1.1 The prototype

The case study was conducted with a software prototype (Figure 5-12) which was a modified and extended version of an existing contouring research software (Steenbakkers et al. 2005). For each GTV contouring task, eight image datasets of a patient, which were in different modalities or acquired in different time during the treatment preparation, were provided. They were: (1) pre-surgery MRI T1-weighted (MRI T1); (2) pre-surgery MRI T1-weighted Contrast Enhanced (MRI T1-CE); (3) pre-surgery MRI FLAIR; (4) radiotherapy treatment planning CT; (5) radiotherapy treatment planning MRI T1; (6) radiotherapy treatment planning MRI T1-CE; (7) radiotherapy treatment planning MRI T2-weighted and (8) radiotherapy treatment planning MRI FLAIR. Prior to the experiment, these eight image datasets were registered to the same coordinate system.

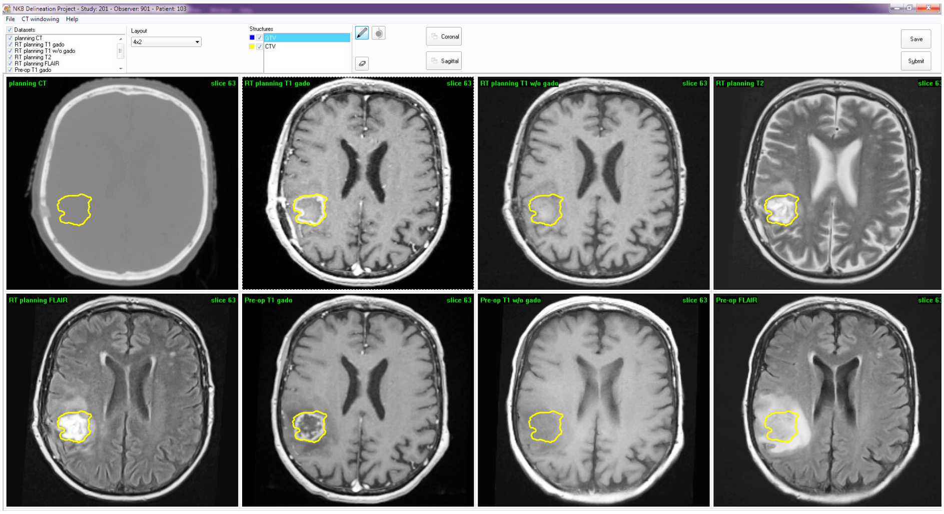


Figure 5-12. The GUI of the software prototype, with a layout representing the same 2D slice of all eight available datasets with a contour overlaid. Physicians typically used layout with 2–3 image datasets.

The GUI of the prototype consisted of the tools area (top region) and images area (middle to bottom region, axial views of all datasets of a patient were provided). Within the prototype, physicians could perform interactions on any of the available eight image datasets. The goal of the GTV contouring task is to contour the visible border of the tumor on all the relevant slices. This was supported by the navigation, data manipulation, and contouring interactions (see Table 5-3). Throughout the task, all interactions with the images were automatically synchronized (i.e., duplicated) to all datasets. For instance, when the physician scrolled to a slice on one of the datasets, the corresponding slices of other visible datasets were presented; if the physician was drawing a contour on one dataset, this contour would immediately appear on all visible datasets at the same location.

Table 5-3. Descriptions of the software interactions available for physicians within the prototype.

Interactions	**Descriptions**
Navigation	
Slice change	Single slice change, switching to the neighboring slice
Scrolling	Series of slice changes, consisting of at least two slice changes
Data manipulation	

Changing datasets	Showing or hiding one or more datasets on the GUI. Changing of the layout of the datasets displayed
Changing active dataset	Switching to a different dataset by mouse interactions
Zooming	Changing the enlargement ratio of a 2D image
Panning	Changing the position of the presented dataset within the GUI
Contouring	
Drawing	Creating, modifying or deleting a contour
Changing drawing tool	Switching from one drawing mode to another
Interpolating	Generating a contour automatically based on neighboring contours

5.2.3.1.2 Participants and the setup of the study

The study was conducted in Department for Radiation Oncology, University Medical Center Freiburg, Germany and Département de Radiothérapie, Institut Claudius-Regaud, Institut Universitaire du Cancer de Toulouse-Oncopole, France. The participants were recruited by senior physicians, resulting in three and five participants from the two hospitals, respectively. The clinical experience of the participants varied: four of them were medical residents, and four were attending oncologists. In each hospital, the study period was a week to accommodate unpredictable clinical tasks. No financial reward was given to the participants.

During the study, the participants were given a task to use the prototype to contour the GTV of GBM which *"consists of the resection cavity and any residual contrast enhancing tumor"*. This was in accordance to the European Organization for Research and Treatment of Cancer (EORTC) guideline, which states that "*GTV delineation should be based on the resection cavity (if present) plus any residual enhancing tumor on contrast-enhanced T1 weighted MRI, without inclusion of peri-tumoural edema*" (Niyazi et al. 2016). Three patient datasets were chosen for the study. These three datasets had been assigned a subjective ranking of difficulty by a senior physician prior to the study: one easy, one medium and one difficult case. Before the study tasks, the participants were given a training session in which they were also allowed to freely explore the software on another sample dataset. Ethical approval for using patient data for research purposes was obtained prior to the study. All physicians participating in the study were informed about the details of the study and signed informed consent forms as well.

In the study, the software prototype was installed and run on a laptop. The display of the laptop was mirrored to a 22-in. monitor, which was the screen size that physicians were familiar with. As input devices, a mouse and a keyboard (with a local language layout) were provided to the physicians. The sequence of GTV contouring tasks of the three patient datasets varied among the participants, in total six possible permutations with no more than two participants for each. Each of the eight physicians contoured three datasets, respectively, resulting in twenty-four cases. The researcher conducting the study was observing the task progress. Necessary help for the software use was provided under the requests of physicians.

5.2.3.1.3 Data analysis methods

The prototype logged mouse and keyboard (i.e., physical) events together with the relevant contextual metadata, e.g., timestamp, the type of the interaction, the duration, the dataset that the physician interacted with, and the slice number, in a log file for later analysis. The log files were then parsed in order to extract the user interactions based on the metadata. For instance, the drawing interaction consisted of a series of mouse-down, mouse-move, and mouse-up events. The periods of no logged physical events were assumed to be cognitive events. These cognitive events, which took place between different interactions, were included within the preceding interaction, resulting in a continuous flow of interactions. For each interaction, relative duration (duration as a percentage of overall task completion duration) was calculated and summed per case as Summed Relative Duration (SRD).

Exploring the details of user interactions allows to bring connections to the reasoning behind (Dou et al. 2009). To enable this, a visual interaction log exploration tool was developed based on JavaScript and D3.js (http://d3js.org) (Aselmaa et al. 2016). The tool enabled interactively exploring interactions of each case as two timeline views: (1) Interaction sequences overview and (2) Interactions on slices overview, as shown in Figure 5-13. The first view, where each interaction was visualized on its own "lane", allowed researchers to identify switches between two consecutive interactions. The second view, where each interaction was displayed in relation to the slice where it occurred, allowed researchers to explore the relations between interactions and their relations to the slices.

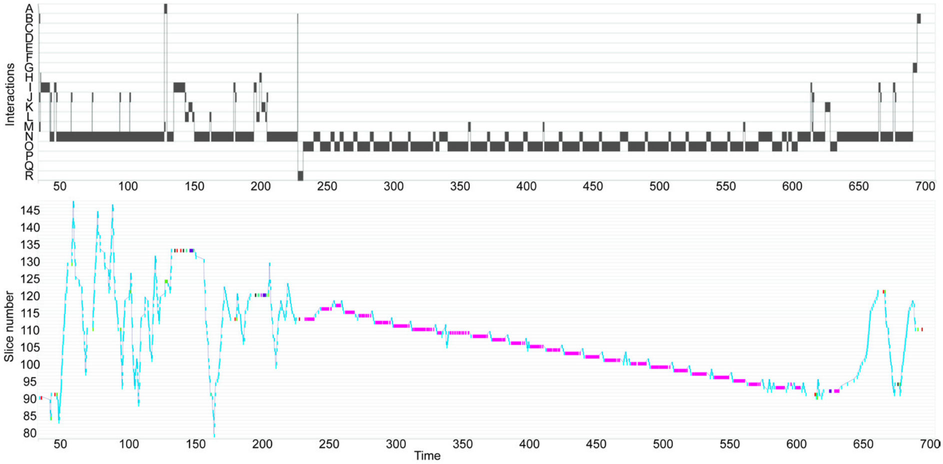


Figure 5-13. Examples of the GTV contouring process timeline as visualized with the tool: (top) interaction sequence overview (interactions labeled from A to R, e.g., N = scrolling, O = contouring); (bottom) interactions on slices overview (different colors represent different types of interactions, e.g., slice change/scrolling is in cyan, and drawing is in magenta).

A navigation interaction or an interaction sequence pattern, representing re-occurring user behavior while using a software solution, carries higher level of meaning than individual interactions. Based on the observed transitions from one interaction to another or from one slice to another in the two visualizations, different types of navigation interactions and interaction sequence patterns were identified. In the process, special attention was paid to situations where the data presented on the GUI changed, as they potentially indicated a change in the sensemaking process. In detail, the labeling of users' interactions was an iterative process as shown in Figure 5-14. The first step was to explore the data for identifying the types of navigation interactions and the interaction sequence patterns and defining the corresponding rules. Then, those rules were programmatically applied to all of the data, and interactions matching the rules were labeled correspondingly. The labeled interactions' data was also presented in a tabular format, so that the correctness of the labeling could be validated. The pattern verification was carried out by two researchers with: (1) the interaction sequence overview; (2) the interactions on slices overview; (3) the tabular labeled interaction data and (4) the rules of different types of navigation interactions and interaction sequence patterns. Each of them individually checked the labeled interaction data and added, corrected or removed the labeling of a possible type of navigation interactions or interaction sequence patterns according to their preferences. Subsequently an inter-rater reliability study was conducted to verify the findings. In the case of disagreements, two researchers went back to previous steps to understand the discrepancy in the data and/or to identify possible new rules. The whole process iterated until a satisfied result was obtained.

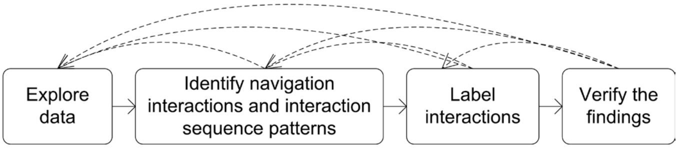


Figure 5-14. The process of identifying types of navigation interactions and interaction sequence patterns.

The periods of task phases were marked for each case based on the occurring interactions. The familiarization phase could be identified as one continuous period, while the action and the evaluation phases were alternating. Familiarization phase was defined as from the beginning of the task until the first contouring interaction. The action phases could be defined mostly based on the contouring interactions. The evaluation phase was typically intertwined with the action phase, consisting of navigation and data manipulation interactions. In most cases, the task ended with a longer period of evaluation.

Each of the navigation interactions and interaction sequence patterns could be associated with a task phase (familiarization, action, or evaluation) based on the primary interactions involved within it and the moment of occurrence in relation to the overall task progress. The duration, occurrence frequency, and slice change count of them were calculated when applicable. In addition, for the interaction sequence pattern, the ratio of the duration of the cognitive events to the duration of the physical events (CE/PE ratio) was calculated when possible (e.g., it was not possible to calculate when no duration was recorded for a physical event). Here the CE/PE ratio 0 indicates only physical events, ratio 1 indicates equal distribution between physical and cognitive events, and the higher the ratio is, the longer the duration of cognitive events is. It is worth mentioning that the CE/PE ratio is limited to the data that a software solution can capture. Thus, for the interactions based on individual mouse events (e.g., left mouse click), the physical events correspond to the speed of the system, rather than the speed of the overall (human) physical interaction time. Nevertheless, the CE/PE ratio gives a relative measure to compare interactions or patterns to each other for their cognitive engagement.

5.2.3.2 Results

The average task completion time was 11 min 26 s (Standard Deviation (SD) = 6 min00 s). Among the total task completion time, the average duration of the familiarization phase was 2 min 6 s (SD = 51 s). The average duration of the action phase, which was calculated as the sum of contouring interactions, was 5 min 47 s (SD = 3 min 47 s). The rest of the time, on average 3 min 33 s (SD = 2 min 00 s), could be accounted for the evaluation phase. The most time consuming individual interactions were drawing (mean SRD 44.4%) and scrolling (mean SRD 39.3%). For the rest of the interactions, the SRD of each was 5% or less.

Based on the visualizations of the interaction sequence overview and the interactions on slices overview (Figure 5-13) of each task, using the data analysis method described in Section 3.1.3, we were able to identify four types of navigation interactions and five types of interaction sequence patterns. Although several iterations were necessary for each case, we found that a high level of agreement between researchers can be achieved in the first iteration. For instance, in six typical cases where four physicians and three patient datasets were engaged, 529 occurrences of navigation interactions and interaction sequence patterns were identified by two researchers in the first iteration. Among them, navigation interactions occurred 141 times (the Cohen's kappa between the results of the two researchers was 0.957, $p < 0.001$) and interaction sequence patterns occurred 388 times (the Cohen's kappa between two researchers was 0.785, $p < 0.001$). Regarding each of the six cases, the Cohen's kappa between two researchers was: 0.901 ($p < 0.001$), 0.891 ($p < 0.001$), 0.933 ($p < 0.001$), 0.837 ($p < 0.001$), 0.901 ($p < 0.001$) and 0.819 ($p < 0.001$). In the following Sections 3.2.1 and 3.2.2, details of those identified navigation interactions and interaction sequence patterns will be presented, respectively.

5.2.3.2.1 Navigation interactions

Navigation interactions (i.e., slice change interactions and scrolling interactions) were time-consuming interactions that represented the thought process of the physician in terms of the 3D navigation. While a single slice change consisted of two sequential events (i.e., navigate to a neighbor slice and perform cognitive actions), a scrolling interaction consisted of multiple navigation-cognition cycles representing a more complex thought process. A single slice change interaction on average lasted for 1211 ms (millisecond, SD = 1093 ms). At the same time, during a scrolling interaction, the average visible time of a slice was 403 ms (SD = 259 ms) during familiarization and 739 ms (SD = 439 ms) during evaluation phase. At the same time, a scrolling interaction involved on average 14.3 slice changes, with a clear difference comparing to the familiarization and the evaluation phases – on average having 28.5 and 10.7 slice changes, respectively.

On the interaction log visualization graphs, it was observed that the physicians' scrolling behaviors varied during different moments of the task. For example, in the beginning of the task they tended to navigate through a wide range of slices, while between contouring interactions they typically navigated in the proximity of a few slices. In order to analyze the variations of different navigation behaviors in relation to the task phases, the navigation interactions were categorized based on the range of the slices they included: the *single slice navigation* involved only one slice change, the *neighbor navigation* involved up to five slices with maximum distance of two slices from the starting one, the *region navigation* involved up to ten slices, and the *long navigation* involved more than ten slices. These four types of *navigation* interactions occurred in total 361, 364, 309, and 278 times for the *single slice*, *neighbor, region,* and *long navigation*, respectively.

On average, the *single slice navigation* lasted for 1.2 s, the *neighbor navigation* for 2.8 s, the *region navigation* for 5.4 s, and the *long navigation* for 12.5 s. For these four types of navigation interactions, the average duration and the average visible time per slice were all less during the familiarization phase than the evaluation phase (Average duration: 1195 ms vs.1212 ms, 1.7 s vs. 2.8 s, 4.2 s vs. 5.6 s, and 11.2 s vs. 13.7 s; Average visible time per slice: 1195 ms vs. 1212 ms, 651 ms vs. 902 ms, 559 ms vs. 672 ms, and 337 ms vs. 449 ms, for *single slice*, *neighbor, region, and long navigation,* respectively as Table 5-4. The *long navigation* represented rapid navigation through the datasets, during which one 2D image slice was visible on average 394 ms. Compared to the *long navigation*, the *region navigation* was slower in terms of the duration of slice being shown; the average time per slice was 656 ms. The *neighbor navigation* was mainly present during the evaluation phase (in total 9 occurrences during familiarization vs. 355 during evaluation). The *neighbor navigation* was slower than the *region navigation* as the average slice visible time was 240 ms longer. It can be assumed that the longer focusing time per slice indicated higher cognitive engagement of physicians. Same as the *neighbor navigation*, the *single slice navigation* was also mainly present during the evaluation phase (31 occurrences during familiarization vs. 330 during evaluation). We also found that generally the less the number of slices involved in a navigation interaction was, the longer the visible time per slice was. Thus, a navigation interaction that involved less slices can be seen cognitively more demanding.

Table 5-4. The identified four types of navigation interactions; s = second, t.p.s = time per slice, ms = millisecond.

Types of Navigation interactions	Description	Task phase	Total count	Mean duration (s)	Mean visible t.p.s (ms)	Mean slice change count
Single slice navigation	Scrolling that involved only one slice change	Familiarization	31	1.2	1195	1
		Evaluation	330	1.2	1212	1
Neighbor navigation	Scrolling that involved up to five slices with maximum distance of two slices from the starting one	Familiarization	9	1.7	651	2.6
		Evaluation	355	2.8	902	3.2
Region navigation	Scrolling that involved up to ten slices	Familiarization	44	4.2	559	7.5
		Evaluation	265	5.6	672	8.9
Long navigation	Scrolling that involved more than ten slices	Familiarization	138	11.2	337	36.9
		Evaluation	140	13.7	449	33.4

In addition, for *long, region* and *neighbor navigations,* it was observed that in some situations they occurred only in one direction. Those single direction navigations could be related to two types of behaviors: jumping over some slices or a systematic evaluation. The first type, jumping over some slices, was encouraged by the presence of the contour interpolation interaction. The interpolation allowed the physicians to contour on a few slices, and then use the interpolation to automatically fill in the "blank" slices. Thus, the "jumping slices" behavior did not have strong relation to the sensemaking process, as it was an extension of a contouring strategy. On the other hand, the second behavior "systematic evaluation" was a sensemaking-intense interaction sequence pattern, during which the consistency of contours in different slices could be evaluated in a continuous way. While engaged in systematic evaluation, physicians spent more time on each slice than they spent on "jumping slices".

Table 5-5. Overview of the identified interaction sequence patterns; s = second, t.p.s. = time per slice, ms - millisecond, CE/PE ratio = cognitive event to physical event ratio.

Interaction sequence pattern	Interactions involved	Description	Task phase	Total count	Mean duration (s)	Mean t.p.s. (ms)	Mean slice change count	Mean CE/PE ratio
Continuous zooming and panning	Zooming; Panning	The physician iteratively zooms and pans the datasets	Familiarization	19	8.9	–	–	1.1
			Evaluation	9	9.8	–	–	0.9
Dataset layout change before active dataset change	Changing datasets; Changing active dataset	The datasets presented on the GUI are changed, and the interaction will continue on a different dataset	Familiarization	103	5.5	–	–	–
			Evaluation	36	2.8	–	–	–
Scrolling on a new dataset	Changing active dataset; Scrolling	The physician switches the dataset that they are scrolling through	Familiarization	148	8.4	405	27.6	10.9
			Evaluation	82	7.3	490	18.9	10.0
Systematic contouring	Drawing; Changing drawing tool; Interpolating;	The physician is continuously drawing on neighboring slices	Evaluation-Action	242	33.3	4356	6.8	1.3

	Slice change; Scrolling							
Scrolling which results in a single slice contouring	Scrolling; Drawing; Changing drawing tool; Interpolating	The physician scrolls through a dataset, followed by a contouring interaction within one slice	Evaluation-Action	133	14.4	–	13.3	3.2

5.2.3.2.2 Interaction sequence patterns

Through the visual analysis of the interaction logs, five interaction sequence patterns were identified as listed in Table 5-5. The descriptive statistics of each of the patterns was calculated in relation to the task phase. The mean time per slice (t.p.s.), and the mean slice change count could be calculated for the patterns that involved navigation on multiple slices. However, for the pattern *scrolling which results in a single slice contouring*, the mean t.p.s. was not calculated since it would not reflect the interactions correctly as the navigation interaction (involving multiple slices) preceded contouring interaction (involving only one slice).

The *continuous zooming and panning* pattern was not a frequently used pattern. In total it appeared 19 times during the familiarization phase and 9 times during the evaluation phase. The *data layout change before active dataset change* pattern appeared more often during the familiarization phase than during the evaluation phase (total 103 vs. 36). The software presented two datasets side by side in the beginning of the task, thus the high frequency of dataset changes could be associated with the needs of inspecting more datasets than what was suggested by the software. *Scrolling on a new dataset* indicated a shift of cognitive focus and also more frequently appeared during the familiarization than during the evaluation phase (total 148 vs. 82). All physicians were engaged in *systematic contouring*, which happened on average 10 times during the task with the average duration of 33.3 s. Both *systematic contouring* and *scrolling which results in single slice contouring* represented interaction sequence patterns that were divided between the action and the evaluation phases.

The average occurrences of the different types of navigation interactions and the interaction sequence patterns were found to be around 87 per task. The identified five interaction sequence patterns covered on average 77% (SD = 7.9%) of the total task duration in the 24 cases as illustrated in Figure 5-15. By including all occurrences of the navigation interactions, the coverage approached 92% (SD = 5.5%). The navigation interactions, which were embedded within the interaction sequence pattern, were on average 27% (SD = 7%) of the total interaction time.

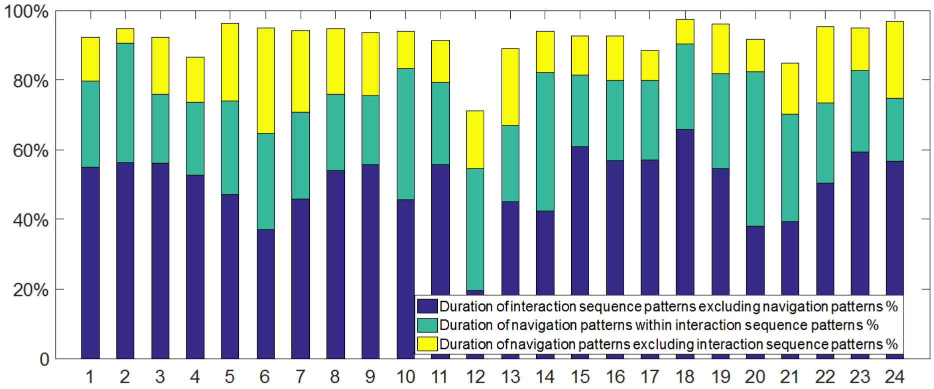


Figure 5-15. Total coverage by the navigation interactions and the interaction sequence patterns in relation to the task completion time (100%) for all 24 cases.

5.2.4 Sensemaking and design insights from the case study

The second step of our proposed approach is to analyze interactions, more specifically the navigation interactions and the interaction sequence patterns, from the perspective of the C-SM model. Each of the identified navigation interactions and interaction sequence patterns involves sensemaking and software interactions to a certain extent. For example, using interaction sequence pattern *dataset layout change before active dataset change* to compare two images side by side for identifying *data or frame gaps* might include few software interactions (e.g., changing data layout) – thus in the use of this interaction sequence pattern, one would be primarily involved in the sensemaking. Another type of interaction sequence pattern could be one that utilizes more heavily the motor skills (e.g., mouse movement, clicking), while cognition is engaged to the extent of deciding on the needed type of interactions and for judging if the goal was achieved, e.g., *systematic contouring*. Thus, identifying the type of the interaction sequence patterns enables identifying potential areas of improvements, for example, for efficiency and/or effectiveness. Table 5-6 presents an overview of the main inferred sensemaking activities and design insights in relation to the task phases and the identified navigation interactions and interaction sequence patterns. This was achieved by positioning each of the navigation interactions and interaction patterns within the C-SM model to gain insights about the sensemaking process and to generate requirements for the interaction design.

Table 5-6. Overview of the main sensemaking inferences and the corresponding design insights from the case study. The sensemaking activities are often interlinked.

Type	**Task phase**	**Inferred sensemaking activity**	**Indication of the sensemaking activity**	**Design insight (category)**
Type of navigation interaction				
Long navigation	Familiarization	Building the initial tumor frame	High number of slices viewed in the beginning of the task	Support effective initial frame creation (1)
		Exploratory information seeking	Extensive data browsing	Support exploring datasets while reducing interactions (2)
	Evaluation	Focused information seeking	Extensive data browsing and relatively slower data exploration (increased cognition).	Support contour evaluation in 3D space (4)
Region navigation	Evaluation	Focused information seeking	Navigating within the proximal data	Support focused/region based inspection of image and/or contour data (3); Support contour evaluation in 3D space (4); Support identifying regions for correction (3)
Neighbor navigation	Evaluation	Focused information seeking	Navigating within the proximal data	Support quick comparison among neighboring slices (4); Support identifying regions for correction (3)
Interaction sequence patterns				
Continuous zooming and panning	Familiarization	Focused information seeking	Increased magnification level. When the magnification level increases, one’s	Reduce time and physical effort (5); Support detecting regions of interest (3)

Type	Task phase	Inferred sensemaking activity	Indication of the sensemaking activity	Design insight (category)
			viewing is more focused (Pantanowitz et al. 2012)	
Dataset layout change before active dataset change	Familiarization	Data/frame gap	New data presented on the GUI in preparation for shifting focus.	Allow user to quickly shift among datasets without additional interactions (2)
Scrolling on a new dataset	Familiarization	Data/frame gap	Shifting focus	Support identifying the relevant datasets for inspection (1)
		Building the frame	Navigating through datasets	Support exploring datasets while reducing interactions (2)
	Evaluation	Data/frame gap	Shifting focus	Support exploring datasets while reducing interactions (2)
Systematic contouring	Action (contouring interactions)	Task analysis	Choices of contouring strategy (precise, rough or none)	Support identifying regions for correction (3)
		Data/frame gap, Building the frame, Preparing the output	Creating and updating contour data. Contouring interaction is an externalization of the (updated) frame	Reduce time and physical effort (5); Support identifying regions for correction (3)
	Evaluation (navigation interactions)	Focused information (gap) seeking	Navigating within the proximal data	Support contour evaluation in 3D space (4)
Scrolling which results in a single slice contouring	Action (contouring interactions)	Preparing the output	Updating existing data	Support identifying regions for correction (3)
	Evaluation (navigation interaction)	Data/frame gap	Updating contour data	Support identifying regions for correction (3)

Data gap = there is not enough information from data; Frame gap = there is not enough knowledge or the mental model is still incomplete.

5.2.4.1 Sensemaking insights

In this section, we attempt to bring connections among the sensemaking activities (as shown in Figure 5-11), the types of navigation interactions and the identified interaction sequence patterns. These conclusions are reached based on knowledge of the context, the software prototype and the meaning of each type of interaction.

5.2.4.1.1 Familiarization phase

Throughout the familiarization phase, we observed that physicians navigated through a number of datasets. The software prototype could display eight available image datasets in various grid layouts. Physicians typically selected two or three datasets to be displayed at once, but there were also physicians who preferred to work with only one dataset, or all eight datasets. Change of the datasets presented on the GUI influenced the sensemaking process, thus the pattern *dataset layout change before active dataset change* was one of the indicators of a *data or frame gap*. The pattern *scrolling on a new dataset* indicated a shift of focus of the dataset physician primarily used, thus it indicated that a *data/frame gap* was found and the frame building process was occurring. It was also found that the *dataset layout change before active dataset change* frequently preceded the *scrolling on a new dataset*, which indicated presence of a gap – the dataset physician needed was not available on the GUI. For example, the physician wanted to see the datasets acquired prior to the surgical intervention to be able to understand where the tumor was before, then he/she compared the acquired information to how it is now for building a hypothesis on the probable extent of the tumor.

The primary type of scrolling during the familiarization phase was the *long navigation*, which occurred approximately five times per case. On average, each *long navigation* led to 36.9 slice changes, during which each slice was visible for 337 ms on average. The *long navigation* during familiarizations enabled browsing through the data and initializing the *initial tumor frame*. Based on the nature of the *long navigation* (rapid exploration of above average number of slices), it can be assumed that it represented the sensemaking activity *exploratory information seeking*, both for *data* and *frame seeking*, resulting in identifying *gaps* and updating *frames* (knowledge update) and/or *data* (data presentation change).

The *continuous zooming and panning* pattern indicated iteratively changing the zooming level and re-positioning (i.e., panning) the 2D image in a preferred way. Increasing the zoom level enabled the physicians to focus on a specific region and to engage in the focused information seeking process. However, it could be assumed that the zoom interaction, immediately followed by the panning interaction, indicated that the zooming functionality on its own was not optimized to the physician' expectations. At the same time, a reduced zoom level could allow the physician have a holistic view of the anatomy (e.g., symmetry between right and left side). As a data manipulation pattern, it influences the sensemaking (new presentation of the data needs to be fitted with the frame) and may result in updating the frame.

5.2.4.1.2 Action phase

The intent for performing the contouring interaction (e.g., *preparing output*) could be seen as an outcome of the sensemaking. While there was a clearly observable transition between the familiarization and the action phases, the transitions between the action and the evaluation phases were fuzzy and more frequent. As a result, physicians had typically more than one contouring episode (i.e., continuous contouring interactions) during the GTV contouring.

The contouring process within a slice consisted of an initial contour creation, (optional) immediate corrections, and (optional) later stage corrections. After the initial contour was created within a slice, two types of immediate corrections could follow: correction for mouse inaccuracy, or for matching the *initial frame* with the contour. For instance, in a line-tracing task it was shown that the mean error with a mouse was 5.8 pixels (Zabramski 2011). Later stage corrections took place after the physician had obtained additional information (i.e., after updating the *specific frame*), often after exploring neighboring slices (i.e., *neighbor navigation*).

Depending on the personal preferences, the specific contouring intention, and the available data, the physicians engaged in different contouring strategies (result from the *task analysis* activity). All physicians were engaged in *systematic contouring* to some extent. Some physicians took a "precise" contouring strategy - they focused on creating a precise contour within a slice before moving to the next slice and often did not make any later stage corrections (see the example in Figure 5-13). Others who preferred a "rough" contouring strategy, often created a rough initial contour first and corrected it later. In some cases, neither of these approaches was followed.

When the physician was following one of these two strategies, there were fewer but longer *systematic contouring* patterns during the case. At the same time, more frequent occurrences of the *scrolling which results in a single slice contouring* pattern indicated the tendency towards a “rough” strategy or no clear strategy.

The *scrolling which results in a single slice contouring* pattern appeared more frequently during the second half of the task. This pattern was potentially an indicator of the *gap* seeking activity. The scrolling portion of this pattern was part of the evaluation phase, while the contouring part was within the action phase. The physician was evaluating the results by scrolling through the data. Once there was a discrepancy identified between the frame and the data, the physician made a correction on the contour. When the correction was done, the physician continued with navigating through the rest of the data.

5.2.4.1.3 Evaluation phase

During the evaluation phase, the *long navigation* may be associated with the focused information seeking activity. For instance, when the physician’s objective was to evaluate the completeness of the contours in 3D, he/she tended to focus on specific areas of the contour. Similarly, the *region navigation* may have represented the focused information seeking activity as well. In this type of navigation interactions, the physician focused on a range of slices, with the aim of evaluating the morphology of the tissue against the contour in order to determine whether there were *data* or *frame gaps*. Sometimes physicians initiated *the scrolling on a new dataset* pattern if the current modality could not offer enough information, and thus the active dataset was changed to the desired modality.

Once a gap was identified, patterns such as *scrolling which results in single slice contouring* or *systematic contouring* were performed to bridge that gap. The *neighbor navigation* occurred typically during *systematic contouring*. Different types of *neighbor navigations* were observed. Examples of them were: viewing one neighboring slice, viewing both neighboring slices, viewing one neighbor and continuing to the other, or viewing a distant neighbor and returning as illustrated in Figure 5-17.

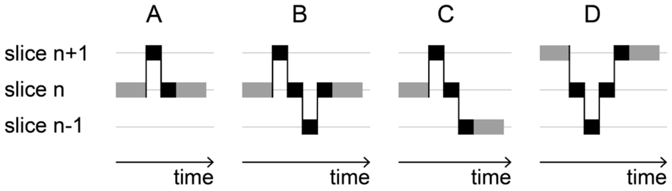


Figure 5-16 Examples of different types of neighbor navigations, each horizontal lane represents a slice. Lighter rectangles indicate the starting and ending slice, while the darker rectangles indicate the change of slice. (A) viewing one neighboring slice; (B) viewing both neighboring slices; (C) viewing one neighbor and continuing to the other and (D) viewing a distant neighbor and returning to the original.

Viewing neighboring slice(s) allowed the physician to re-frame through the visual comparison of the current contour with the neighboring contours/images. It enabled the physician to build a detailed frame of the morphology of tissues within a narrow region and thus helped him/her to gain a better understanding of the tissue dynamics. The two distinct types of comparisons were: (1) comparison of neighboring contour(s); and (2) comparison of neighboring 2D image slice(s). Comparing contours allowed the physician to (re-)evaluate a prior decision, and to determine whether to follow the same principle or modify the contour on the previous slice(s). Comparing 2D images allowed physicians to fill their data gaps, for example, when information in current slice was not definitive, but based on information in neighboring slices, a more concrete assumption could be made. The perceived and projected data was then fitted into the frame, resulting in an updated frame.

5.2.4.2 Design insights

Insights of the sensemaking process help designers identify opportunities for possible improvements to increase the sensemaking support in software design. In this section, we first elaborate on how to utilize the C-SM model to generate design insights. Using this method, we summarize the design suggestions obtained from the case study.

5.2.4.2.1 Using the C-SM model to generate design insights

The main focus of using the C-SM model for generating design insights is to make the design more effective and efficient regarding the sensemaking process. Here effective sensemaking means that one is able to identify the right *frame(s)*, and the corresponding *gaps* between the *data* and those *frames*. Improving effectiveness means supporting the *framing* loops while enabling the right software interactions. Efficient sensemaking, similarly to efficient use of software, means that the goal is reached with least effort and time.

Those primary indicators contain the duration, frequency, and distribution between the underlying physical and cognitive events of the involved interactions. For instance, long-durational interaction sequence patterns involving intense user interactions could be associated with decreased efficiency and increased physical workload. Numerous loops of the same type of interactions could indicate ineffective design and/or lacking data presentation, which demand frequent sensemaking-interaction loops in addition to potential inefficient interaction issues. Interaction sequence patterns with lower cognitive involvements result in short interaction loops consisting of mostly physical events. Improving or eliminating (i.e., automation) these types of interactions can be considered for improving the efficiency. Interactions involving higher levels of cognition are more suitable subjects for effectiveness improvements.

While the duration and frequency of interactions are easily measurable, the level of cognitive involvement is difficult to quantify. We propose to use the CE/PE ratio (as seen in Table 5-5) as an indicator of the cognitive involvement during interaction sequence patterns. The CE/PE ratio compares the cognitive involvement to physical activities in different types of interactions (patterns), thus enables building assumptions on which types of interactions (navigation interactions or interaction sequence patterns) are more cognitively demanding.

5.2.4.2.2 Design insights from the case study

The identified four types of navigation interactions and five types of interaction sequence patterns were positioned within the C-SM model according to the types of interactions they included and during which phase they occurred. Then, each of them was analyzed regarding the task phase and the involved sensemaking activities. Example questions that were asked during this analysis were: “What kind of *data-frame gaps* are present?”; “Which sensemaking activities may enable the physician to identify the *gaps*?” and “How could (other) interactions, or different GUI elements, support bridging the *gap*?”. Based on the analysis of each pattern in relation to the sensemaking process, the key design requirements for supporting sensemaking are generated. Table 5-7 highlights the primary indicators for sensemaking support improvements and their types. The main design requirements for supporting sensemaking can be categorized to the following five areas: (1) to enable effective initial frame development, e.g., support identifying the relevant datasets for inspection; (2) to support intuitive navigation within and between datasets, e.g., support exploring datasets while reducing interactions, allow the user to quickly shift among datasets; (3) to support detecting regions of interests; (4) to enable additional methods for contour evaluation e.g., 3D evaluation, neighbor comparison and (5) to improve the general efficiency by reducing time and physical efforts. Those requirements are summarized as the final column in Table 5-6, corresponding to the sensemaking activities which they support, respectively.

Table 5-7. Overview of potential indicators for improvements regarding effectiveness and efficiency.

Indicator for possible improvements	**Type(s) of improvements**
Long duration of cognitive events	Effectiveness
High frequency of cognitive events	Effectiveness
High cognitive involvement	Effectiveness
Use of external tools/materials	Effectiveness
Long duration	Effectiveness and Efficiency
Alternating cognitive events with physical events	Effectiveness and Efficiency
Long overall duration	Efficiency
Low cognitive involvement (short durations of cognitive events)	Efficiency
Repetition of the same type of interaction	Efficiency

Using those design requirements, we are able to propose possible improvements to support sensemaking in the software design. For instance, the *long navigation* during familiarization phase is about building an *initial frame*, which bridges the *gap* between the previously unknown data and the *general tumor frame*. This is achieved through *exploratory information seeking*. This information seeking was supported by navigation interactions within the study prototype. As an alternative, an “autoplay” function could be designed for *exploratory information seeking* that is already optimized in terms of data range involved and the speed of slice changes. Furthermore, since we observed oscillating scrolling behavior during the *long navigation* before physicians focused on a slice, the "autoplay" function could simulate this as well. However, fully automating this type of scrolling might restrict the needed interactions of the physician, thus the "autoplay" function could be triggered by the physician after opening the patient dataset, while still allowing manual interaction afterwards.

Some requirements were identified from multiple patterns, for example, the requirement "Support identifying regions for correction". The design improvements for this requirement can be providing medical knowledge and/or possible technical supports. From the medical perspective, improving the interface design to incorporate medical knowledge of what regions should (or should not) be included within the GTV contour may guide physicians in the process. A simple solution could be presenting a checklist, which the physician has to revise prior to completing the task. However, such solution may decrease the overall efficiency. From the technical perspective, a more complex solution could be achieved by embedding medical knowledge in computational algorithms to provide immediate feedback. For instance, developing a function that is able to evaluate the 3D consistency of the shape by comparing a 2D contour to other contours on the neighboring slices.

5.2.5 Discussion

5.2.5.1 The case study

Analyzing interaction logs in order to comprehend the underlying reasoning is a growing field of interest. Through examining the interactions, it is possible to identify 60–79% of strategies/methods (Dou et al. 2009). We limited our research to analyzing interactions based on the visual inspection of the software interaction timelines. In our case study, with the limited number of cases and interactions, visual inspection was found sufficient and we were able to identify five main interaction sequence patterns covering on average 77% of the overall task duration. In combination with the different types of navigation interactions, the coverage reached 92%. Meanwhile, automated pattern mining solutions could give additional benefits when the sample size is larger. Compared to field studies, the pattern mining approach has limited effects in identifying usability issues (Jorritsma et al. 2016). However, we have shown that identifying patterns is beneficial for generating deep insights on how a software solution is used and about the underlying sensemaking process.

Within our study the aim was to identify main navigation interactions and interaction sequence patterns and infer their possible relations to sensemaking activities. More detailed interaction sequence patterns could be developed (e.g., depending on the case/tumor size) to enable even more in-depth analysis. At the same time, it is important to acknowledge that the types of navigation interactions and interaction sequence patterns strongly depend on the task and the context, thus context specific pattern rules are often needed.

In a non-computer aided solution, the contouring task requires the physician to draw the visually seen borders on the 2D images slice by slice. Such approach is time consuming, and thus research on semi-automatic and fully automatic segmentation of the tumor is being conducted and some promising results have been achieved. However, a general conclusion on their accuracies is difficult to make, as they have been evaluated on individual cases (Bauer et al. 2013). Furthermore, most of those algorithms still require human involvement (Heckel et al. 2013). With the development of automatic segmentation algorithms, it can be foreseen that the GTV contouring task will gradually change to be a task of evaluating and correcting the outcome from computational results (contours). As such, it is crucial to increase the support for sensemaking regarding comprehending the generated initial contours, identifying regions for correction and enabling new ways for evaluating the contours. Furthermore, intelligent tools for contour corrections will be needed. For example computer-aided contouring tools, that perform immediate adjustments to the drawn contours based on the information available on the medical image(s) have shown promise in decreasing the overall interaction time (Fitton et al. 2011).

Although computational algorithms seem promising, in our case study, we were able to bring out that drawing interactions are only a part of the overall process (on average accounting for 45% of the total interaction time). The efforts of developing computational algorithms for generating contours (Zhao and Xie 2013) as such, can only automate part of the overall task. Therefore, for a better software design, more efforts are needed to support all phases of the task, by integrating computational algorithms as well as supporting the sensemaking process.

5.2.5.2 The two-step approach

We proposed a two-step approach for incorporating sensemaking in order to identify additional design inputs. The first step was to model the sensemaking in context, where the C-SM model was developed based on the generalized sensemaking model and by incorporating the knowledge of the task phases and the needed software interactions. However, in reality, sensemaking is a complex phenomenon and an in-depth understanding of the details of the sensemaking process are required in order to contextualize the sensemaking process. Thus, our proposed approach is to be seen as a supporting tool that enables designers to connect the software interactions with a sensemaking theory during the design process, but not as a replacement of existing sensemaking theories. It is worthwhile to mention that besides literature research, we used observational research to acquire contextual knowledge. Though considerable time and efforts were spent on the research, it offered rich contextual information and made the task tangible to the HIS designer. In HIS design, many tasks are in very specific contexts and highly complicated. Though expensive, we recommend HIS designers using observational research to get acquainted with the context and generate a qualitative description.

The second step of our approach is to analyze the software interactions through the C-SM model. We suggest analyzing types of navigation interaction and interaction sequence patterns instead of individual interactions. Compared to an individual interaction, a pattern represents a significant software use behavior, thus incorporates more high-level and contextual information. Furthermore, navigation interactions and interaction sequence patterns not only give valuable insights of the sensemaking activities, but also enable identifying shifts between task phases.

Though our approach is only demonstrated on a case study of tumor contouring, it could be applied in other data-driven sensemaking contexts as well. First, the modeling step could be adapted to different contexts. For developing the C-SM model, the generalized sensemaking model is sufficient as it is not related to one specific context or sensemaking theory. During data-driven sensemaking, both data and frame(s) are present, where the frames represent the sensemaker's knowledge and experiences of the task and the context. The identified three primary phases of the tumor contouring task could also be generalized as the exploration phase, the action phase, and the verification phase, thus representing the main phases of any problem solving (the action phase is implicit) (Lindsay and Norman 1972). Last, our C-SM model incorporated software interactions relevant for the case study. It is able to associate detailed interactions with sensemaking activities and thus reveal sensemaking activities in an objective manner. Given a different task and its context, by applying the proposed approach, an adapted C-SM model incorporating the relevant software interactions can be developed.

5.2.5.3 Limitations

Within our observational research and case studies, no verbal (e.g., think aloud) protocols were used. Such methods could bring valuable insights and could allow better connection building between user interaction with the software and the sensemaking process. It is expected that a (retrospective) think aloud study could be beneficial in similar cases (van den Haak et al. 2003).

In the analysis of interaction, currently we use the CE/PE ratio as an indicator regarding cognitive involvement during interaction sequence patterns. Though effective, it only can describe the cognitive activities as a whole. An in-depth analysis of those cognitive involvements may reveal more details of the sensemaking process and activities. As part of the future work, we plan to introduce eye-tracking in the experiment in order to discover more details in the sensemaking process.

5.2.5.4 Conclusion

In this section, we proposed a two-step approach for incorporating sensemaking into HIS software design in order to generate design insights. The first step, modeling sensemaking in context, enables designers to describe the position of sensemaking within a task process in relation to the GUI and interactions between the user and the software solution. The second step, in-depth analysis of software interactions (patterns) in relation to the C-SM model enables designers to identify possible improvements of detail interactions regarding both effectiveness and efficiency, which can be highlighted as new design requirements to support sensemaking.

To demonstrate the effectiveness of this two-step approach, we conducted a case study of the tumor contouring task for radiotherapy planning. Within the C-SM model of this task, we described: (1) the three main phases of the task: familiarization, action and evaluation; and (2) sensemaking in relation to the primary software interactions, e.g., navigation, data manipulation, contouring, etc. Through the analysis of the interaction logs of twenty-four cases, we identified four types of navigation interactions and five interaction sequence patterns. Based on the analysis of each navigation interaction and interaction sequence pattern, we discovered five main areas of improvements that may increase the support of sensemaking in the process: (1) to enable effective initial frame development; (2) to support intuitive navigation within and between datasets; (3) to support detecting regions of interests; (4) to enable additional methods for contour evaluation and (5) to improve the general efficiency by reducing time and physical efforts. Based on the outcomes of the case study, it is concluded that the proposed two-step approach has proved to be beneficial for gaining detailed insights of the sensemaking process and deriving design requirements that support sensemaking.

5.3 User interaction visualization

Through the synthesis of gathered user research data, interaction designers are able to generate design proposals. Logging user interactions with a software provides a rich set of data that can give further insights into users' behavior. We present a case study of visualizing interactions log files of the manual tumor contouring task. We identify two types of visualizations needed for comprehending the tumor contouring process. Based on these visualizations, designer was able to gain a holistic view of the process, detailed understanding of the different phases of the task, and identify re-occurring interaction patterns.

5.3.1 Introduction

Design synthesis is an abductive sensemaking process of manipulating, organizing, pruning and filtering previously gathered data regarding users' behaviors, needs and motivations, in an effort to generate design proposals (Kolko 2010). It is a decisive point in the design process, meanwhile it is also rather complicated and exhausting process due to the large variety and amounts of data (Gumienny et al. 2011).

In the design process, a variety of data is collected regarding the users, by utilizing various user research techniques and methods (e.g., interviews, observations, video recording). In addition, there has been increased interest in capturing human computer interaction events (such as mouse clicks and key strokes) in a log file (e.g., (Guo et al. 2016), (Jorritsma et al. 2016)). Such interaction log files can be a rich source of data for comprehending the fuzzy user behavior. However, comprehending the raw (typically textual) log files is very challenging for designers due to the limitations of human short-term memory (Proctor and Wu 2007). At the same time, design synthesis (a task that has fuzzy task clarity) based on digital data, can benefit from visualization (Sedlmair et al. 2012).

In this work, we present two visualizations to help the designer to better comprehend information from interaction log files, based on a case study of manual brain tumor contouring on 3D medical image datasets. Based on the outcomes of this case study, we also highlight requirements for an interactive visualization tool that enables grasping user interactions during medical image contouring.

5.3.2 Case study

Tumor contouring is a complicated clinical task, completed through the use of a dedicated treatment planning software. In order to comprehend the behavior of users and to identify requirements for an improved design, we conducted a contouring study where user interactions were captured in a log file. The studied task consisted of two sequential sub-tasks as in a typical clinical routine (Burnet 2004): contouring the Gross Tumor Volume (GTV), and correcting the automatically generated Clinical Target Volume (CTV). In total, eight physicians of varying level of experience joined the study, and each of them contoured three datasets. This resulted in total of 24 cases. Ethical approval for the use of patient data was obtained prior to the study.

The log files created during the study consisted of timestamped low-level events (e.g., mouse-click, mouse-drag, etc.) that were clustered into meaningful software interactions based on the available meta-data (e.g., a cluster of mouse-drag events could be "drawing", "zooming", or "panning" interaction). Two data visualizations were developed using the D3.js library to visualize: 1) the sequence of interactions (*Viz-Seq*) and 2) the interactions on different 2D slices (*Viz-Slice*) as Figure 1.

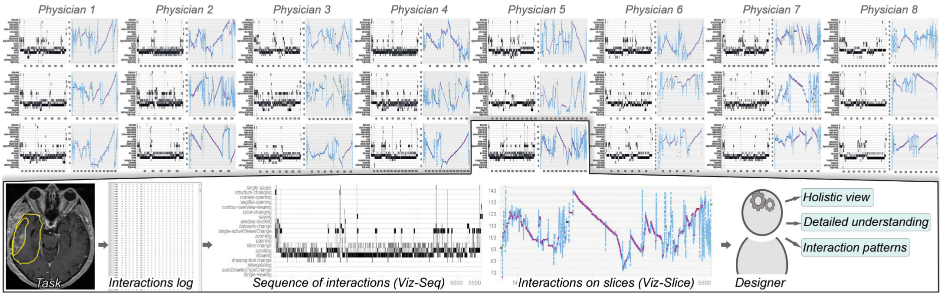


Figure 5-17 Visualizations from the case study of the manual tumor contouring task, with a detailed example.

The *Viz-Seq* view was developed to indicate the shifts between interactions. Each type of interactions (in total 19) displayed in the *Viz-Seq* view had a dedicated horizontal lane, within which each occurrence was represented by a rectangle. The transitions between different interactions were indicated by connecting lines, for easier tracing of the interaction sequences. The *Viz-Slice* view was developed to show the occurring interactions on each slice. In the *Viz-Slice* view, each horizontal lane represented a 2D slice within the 3D dataset, and each interaction was represented by a color-filled rectangle. The designer also had the possibility to select a subset of interactions to be visualized to eliminate visual noise. In both views the width of the rectangle indicated the duration of the interaction. In addition, the designer was able zoom in to a part of the visualization by dynamically changing the time range with a slider. Furthermore, the designer could quickly switch between the different cases.

5.3.3 Synthesis results

Based on the visualizations, the designer was able to explore each of the 24 cases both in a holistic and detailed way. The visualizations helped understanding the task process quicker than reviewing video recordings would have been. At the same time, it also allowed visual comparison of the task process (complete or a portion) between users. Furthermore, it was possible to identify reoccurring interaction patterns that gave further insights into typical software use. Below, we will highlight few of the findings.

Based on previous observational studies, we were able to qualitatively conclude that the tumor contouring task consists of three phases: familiarization, action and evaluation. However, the transitions between these phases and detailed interactions during them were not clear. Using the visualizations it was possible to build an in-depth understanding of the three phases. For example, using *Viz-Seq* it was possible to identify that shifting between datasets not only took place at the beginning of the task (familiarization phase), but also at the end of the task (evaluation phase). The action phase, which mainly consists of focused contouring interactions, was clearly identifiable based on both of the visualizations.

Three types of contouring strategies were identified: 1) the physician created a *precise* contour on a slice before moving on to the next slice; 2) the physician created *rough* contours on multiple slices first, and then iteratively revising the contours in each slice until a satisfied result; and 3) a *mixed* approach that combined both strategies at different moments.

Four scrolling patterns and five interaction sequence patterns were identified. These patterns provided a deeper understanding of user interactions in different phases and helped building hypotheses of underlying reasoning processes. For instance, *rapid scrolling* at the beginning of the task could be associated with getting an initial understanding of the case, while the same pattern at the end of the task could be associated with rapid evaluation of the contours.

5.3.4 Reflections

Two types of visualizations of the interaction log files are required to comprehend the contouring process: one to support the exploration of interaction sequences; and one to identify interactions in relation to the third dimension of the image dataset. In a setting where video is recorded in addition to logging, it could be beneficial to extend the timeline views by embedding the video recording synchronously.

The vertical order of interaction lanes within *Viz-Seq* was predefined - the least occurring ones first, and frequently interchanging ones (e.g., *slice change* and *drawing*) following each-other. This guaranteed visual consistency and eased visual comparison between cases. As an improvement, using color to differentiate sub-types of an interaction could be beneficial (e.g., *drawing* a new line, *drawing* to edit, or *drawing* to delete).

Log files are commonly analyzed for extracting usability measures, identifying usage patterns, inferring knowledge or expertise (Guzdial et al. 1994). Our focus was on better supporting inferring knowledge of users and thus we visualized each case separately. Another approach could be to visualize multiple cases at once (e.g., (Malý and Slavík 2007)).

5.3.5 Conclusions

In this section, we presented two timeline visualizations of log files to explore user interactions based on a case study of tumor contouring: one for visualizing sequences of interactions (*Viz-Seq*), and one for visualizing interactions in relation to the third dimension of the medical image datasets (*Viz-Slice*). Both visualizations allowed the interaction designer to explore the task process in a holistic view as well as in details. Based on the visualizations it was possible to better understand the transition between the phases of the task, as well as the occurring interactions during these phases. Furthermore, the visualizations helped identifying contouring strategies and main interaction patterns. Results from the case study indicate that (*Viz-Seq*) and (*Viz-Slice*) timelines support design synthesis.

Chapter 6

Influence of automation

This chapter is based on:

Aselmaa A, van Herk M, Song Y, Goossens RHM, Laprie A (2017) The influence of automation on tumor contouring. Cognition, Technology & Work. doi: 10.1007/s10111-017-0436-0

Automation helps in improving efficiency while performing tasks. As such, use of automation is inevitable in a contouring software. This chapter presents results from a study to evaluate the influence of automation the outcomes and the process of the tumor contouring task.

6.1 Influence of automation

Automatic contouring tools, such as the interpolation tool, are increasingly being used in the tumor contouring task for radiotherapy. However, automatically generated outputs may influence the reasoning process of physicians. This study evaluates those influences on the resulting contours and the contouring process.

6.1.1 Background

Radiotherapy is one of the most effective methods for the treatment of cancer (Njeh 2008) with an estimate of 52% of cancer patients benefitting from it (Delaney et al. 2005). With the aging population, cancer incidence and mortality are expected to increase (Yancik and Ries 2004). There is an increasing need to optimize the radiotherapy workflows as well as to automate different (parts of) tasks in order to improve the efficiency of radiotherapy and to reduce the workload of the physicians (e.g.,(Olsen et al. 2014; Kirrmann et al. 2015; Winkel et al. 2016)).

One of the tasks within radiotherapy planning where automation has been increasingly introduced is the contouring task. During this task, the tumor and the surrounding organs-at-risk are contoured on the medical images by a physician (Vieira et al. 2016). Manual contouring can be tedious and time-consuming (Dowsett et al. 1992; Vorwerk et al. 2014), and introducing automatic contouring tools (i.e., contouring with the support of automation) is generally expected to decrease the task duration (Lim and Leech 2016). However, automation may influence physicians' decision-making process, i.e., introduce bias. Automation bias is the phenomenon that appears when the automatically generated decision aids are used as a replacement for a more vigilant system monitoring or decision-making (Skitka et al. 1999). This may result in decisions that are strongly guided by those automatically generated advices (Parasuraman and Manzey 2010). Regarding tumor contouring, automation bias may result in the errors of omission and the errors of commission (Skitka et al. 1999). Here, the errors of omission indicate that the automatic contour did not include all the relevant regions, but was still accepted by the physician. The errors of commission, at the same time, would mean that an automatically suggested and accepted tumor contour included also healthy tissue. These errors could lead to missing tumorous tissue during radiation, or irradiating healthy tissue unnecessarily. Therefore, automation bias should be taken into account when introducing automatic contouring tools to radiotherapy planning software (Wesley and Dau 2017).

Using a relatively basic automatic contouring tool, named between slice interpolation, this chapter aims at studying the influences of an automatically generated initial contour on the resulting contours and the contouring process. Three aspects will be explored regarding this topic: (1) The variations among the contours created by physicians with and without the automation tool, as previous research indicated that variability among physicians in manual contouring is a large concern in radiotherapy (van Herk 2004; Fitton et al. 2011); (2) The duration of the contouring task, i.e., the efficiency; and (3) Changes in the contouring workflow introduced by using the automation tool.

6.1.2 Contouring in radiotherapy planning

The planning of radiotherapy involves a number of clinicians and tasks (Vieira et al. 2016). Once radiotherapy is suggested based on the diagnosis and is discussed with the patient, the necessary data for the treatment planning, such as medical image datasets of different modalities (Batumalai et al. 2016), are acquired. Those images may consist of Computed Tomography (CT) images, various sequences of Magnetic Resonance Imaging (MRI) images, and/or Positron Emission Tomography (PET) images, depending on the type of the tumor (Batumalai et al. 2016). All acquired medical image datasets are then co-registered, i.e., aligned to the same coordinate space for inclusion in the planning process (Weersink 2016).

The image co-registration step is then followed by the contouring task, during which the various treatment volumes, i.e., the tumor, as well as the surrounding healthy tissues are contoured by a physician(s) (Vieira et al. 2016). One of the axioms of radiotherapy is to maximize the prescribed radiation dose to the tumor while sparing surrounding organs at risk (Burnet 2004). For this, accurately identifying the location and the shape of the tumor is a prerequisite. This is especially true, as with the technological advancements in image-guided radiotherapy, it is possible to precisely deliver the radiation to complicated 3D volumes (Nutting et al. 2000; Xing et al. 2006).

Different types of volumes are used for the treatment planning as recommended by the International Commission on Radiation Units and Measurements (ICRU) in report 62 (ICRU 1999). The Gross Tumor Volume (GTV), which represents the visible (on medical image datasets) and/or palpable tumor, is the basis of other relevant tumor volumes, such as the Clinical Target Volume (CTV) (Burnet 2004). During the contouring process of the GTV, medical image datasets are presented on the computer screen as 2D images, each of them representing a section of the human body (i.e., “slice”). The physician then draws the visually seen borders of the tumor on a number of slices (Dowsett et al. 1992), resulting in a set of 2D contours representing the 3D volume of the GTV. Once all the relevant contours are created, different 3D volumes are constructed, e.g., by lofting those contours. Radiation dose is then planned and validated (e.g., (Winkel et al. 2016)) based on the dose constraints on these volumes. Among different contouring tasks, the GTV contouring task is especially important in radiotherapy planning since GTV is the basis for defining other volumes for the treatment planning and consequently, uncertainties in this step may introduce a systematic error for the complete treatment planning (van Herk 2004).

The GTV contouring task can be divided into three phases: familiarization, contouring (i.e., action), and evaluation (Aselmaa et al. 2017). Prior to creating any contour, the physician usually spends some amount of time exploring the information presented on the graphical user interface getting familiar with the data. The contouring action itself can be further divided into creating an initial contour(s) and correcting this contour(s), either immediately or later. Then, the contour(s) is iteratively evaluated and modified as needed throughout the contouring process. For example, a physician may first create the initial contours for a few neighboring slices and then continue with iteratively modifying these contours or creating contours on further slices.

Contouring without any computational support can be lengthy and tedious as it requires drawing the visually seen borders of the tumor on all intersecting slices (Dowsett et al. 1992; Vorwerk et al. 2014). In the past decades, extensive research has been conducted and various automatic contouring tools (i.e., segmentation methods) have been developed (Olabarriaga and Smeulders 2001). Some of these tools have been gradually introduced in commercial radiotherapy planning software solutions (Sykes 2014). The expected benefits of using automatic tools are the reduction of the overall amount of time taken to draw the contours, and potentially also increased reproducibility of the contours (i.e., reduced inter-observer variation).

Contouring tools can be categorized as fully automatic, semi-automatic or fully manual (Heckel et al. 2013) based on the intended level of involvement of the physician and computation. Fully automatic contouring is potentially the most time efficient as it is designed to have little involvement of the physician. However, fully automatic contouring methods have shown limited success and often extensive post-processing is needed (Bauer et al. 2013; Sykes 2014). Automation may be introduced in different levels for semi-automatic methods: from automatically generated 3D volumes based on a few 2D contours (e.g., based on the foreground and background seeds (Dolz et al. 2016)) to computationally adjusting the contour while it is being drawn by the physician (e.g., live-wire tool (Barrett and Mortensen 1997)). And different levels of automation may pose different influences (Bravo & Ostos 2017) on the physician's decision-making process. Among different semi-automatic methods, a commonly used category of tools in commercial software solutions is the contour interpolation (e.g., shape based interpolation) (Prabhakar et al. 2011). One such tool is the between slice interpolation which generates a 2D contour based on the nearest contours on the inferior and superior slices, and the physician is expected to make corrections until reaching a satisfactory result. The advantage of such a semi-automatic method is that it accelerates the contouring process by combining the power of computing and human expertise for the initial contours, based on the assumption of the continuity of the tumor shape while allowing physicians to control the outcomes.

Physicians play a central role in steering and correcting the outcomes of the contouring task (Heckel et al. 2013). However, their cognition can be influenced by those automatically generated or corrected contours, especially as there is no gold standard in GTV contouring (Weiss and Hess 2003, Aselmaa et al. 2017). A higher level of automation can introduce a higher level of bias (Manzey et al. 2012). On the other side, lower level of automation, which has a higher level of human involvement, may have a smaller gap between physicians' cognition and the data, thus the influence of the automation can be expected to be smaller. The contouring task is an iterative process during which contours are being inspected multiple times. Therefore, it is expected that the gap narrows even further in this iterative process. However, literature study did not reveal to what extent such or similar interpolation may influence the physicians' decision-making process. The questions about the clinical relevance of such an automation bias and its effects on the inter-observer variation also remain to be answered.

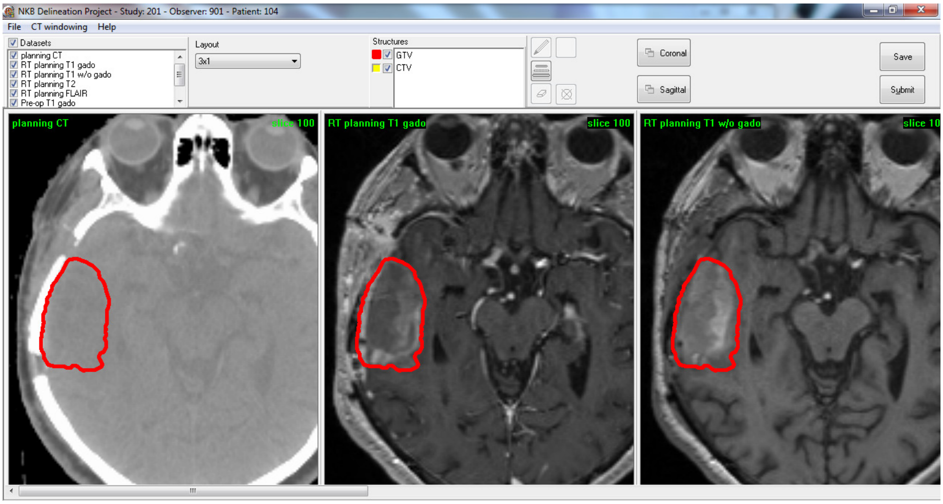
6.1.3 Study methods

6.1.3.1 Study setup

To evaluate the influence of using the between slice interpolation tool on the resulting GTV contours and on the contouring process, a GTV contouring study was conducted in the radiotherapy department of Institut Claudius-Regaud, Institut Universitaire du Cancer de Toulouse-Oncopole, Toulouse, France with five physicians (three medical residents, two attending physicians) over the period of five days. The investigated task was the GTV contouring of the Glioblastoma Multiforme (GBM) tumor, a common type of primary brain tumor (Behin et al. 2003). Four patient cases (a sample case, C-1, C-2, and C-3) were used in the study. Subjective rating of the case difficulties (easy, medium difficulty or difficult) was given by an experienced physician independently from the present study.

Similar to the clinical practice, eight image datasets were made available for the physicians for each case. Those eight images datasets were: radiotherapy planning CT, radiotherapy planning MRI T1-weighted with contrast enhancement, radiotherapy planning MRI T1-weighted without contrast enhancement, radiotherapy planning MRI T2-weighted, radiotherapy planning MRI FLAIR, pre-surgery MRI T1-weighted with contract enhancement, pre-surgery MRI T1-weighted without contrast enhancement, and pre-surgery MRI FLAIR. Prior to conducting the study, the MRI datasets were co-registered to the radiotherapy planning CT coordinate system. The distance between any two consecutive axial slices was 2.5 mm in the case C-1, and 1.25 mm for the other two cases.

The study was conducted using a modified and extended version of a research contouring software (Steenbakkers et al. 2005)(Figure 6-1). The software allowed manual contouring (i.e., using the freehand and/or the nudge tools) and *between slice interpolation* (i.e. using the interpolation tool) on any of the axial slices of any of the available datasets displayed on the computer screen. Using the interpolation tool, a contour could be generated on the displayed axial slice based on the contours on the nearest neighboring slices via linear interpolation where the point correspondences were obtained using a radial coordinate system. Within this study the interpolation tool was used only for creating the initial contour, i.e., the method was only available when there was no existing contour on the slice. The interpolation tool was not available for correcting an existing contour, neither for the first and last slices, as the interpolation relied on the information of the neighboring contours. For the rest, to guarantee the "natural" performance of physicians as it is in a clinical setting, physicians were free to choose either the manual or the interpolation tool to create the initial 2D contours. We expected that such a "randomized" setup will minimize the cognitive and psychological difference in the selection of methods. Meanwhile, the software recorded all user interactions during the task into a log file together with timestamps.



NKB Delineation Project - Study: 201 - Observer: 901 - Patient: 104

☑ Datasets
☑ planning CT
☑ RT planning T1 gado
☑ RT planning T1 w/o gado
☑ RT planning T2
☑ RT planning FLAIR
☑ Pre-op T1 gado

Layout: 3x1

Structures
🟥 ☑ GTV
🟨 ☑ CTV

Figure 6-1 A screenshot of the software prototype used in the study. The contouring tools available in the study are in the top part of the graphical user interface. The image datasets are presented in the middle-bottom region. An illustrative 2D contour of the GTV is shown in red color, overlaid on the radiotherapy planning CT (left image), radiotherapy planning MRI T1-weighted with contrast enhancement (middle image), and radiotherapy planning MRI T1-weighted (right image)

In the beginning of the task, the physician was introduced to the software and a brief training was given with the sample case. In addition, the physician was allowed to explore the software further as they felt necessary. Then, the physician was asked to contour the GTV of the rest of the three GBM cases. Following the treatment protocol (Stupp et al. 2005), the GTV was instructed as "consisting of the resection cavity and any residual contrast-enhancing tumor". The order in which these three cases were presented varied among physicians in order to distribute the impact of learning effects. The researcher was observing the task during the study and was available for assistance with the use of the software at request.

After finishing each GTV contouring task, each physician filled the NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland 1988), which was used for assessing the mental workload of physicians based on the subjective rating on six aspects: the physical demand, the mental demand, the temporal demand, the performance, the effort, and the frustration. The original NASA-TLX consists of two parts: rating each aspect, and comparing them pairwise based on their perceived importance. However, it has been shown that the unweighted and the weighted ratings have a high correlation (Noyes and Bruneau 2007). In this study, the outcome of the NASA-TLX was calculated based on the unweighted ratings.

6.1.3.2 Data inclusion and analysis

Figure 6-2 illustrates different steps in the data inclusion and analysis process. In the proposed research, contouring was made possible on axial slices only. The slices towards the superior and inferior boundary of the tumor typically have a larger level of variation than the central slices. For instance, given a boundary slice, it was often that not all physicians contoured, i.e., not all physicians agreed that there was tumorous tissue. Such a cognitive difference often leads to large deviations among the boundary slices. The influence of automation, if any, was expected to be smaller than other influences. Therefore, two criteria were applied to eliminate “boundary slices”: (1) the slices on which not all physicians contoured ($n < 5$) were excluded from the analysis; (2) for the remaining slices, the mean enclosed areas of each contour ($Mean_{area}$) and the standard deviation (SD_{area}) among them were calculated for each slice over the observers. Then, the coefficient of variation (i.e., relative standard deviation, CV_{area}) within each slice was calculated as the ratio between SD_{area} and $Mean_{area}$. Contours that would be included in the further analysis were defined by its CV_{area} being less than the mean of CV_{area} + 1 SD_{area} of the given case.

Then contours on the included slices were categorized as being *manual* or *interpolated* based on whether interpolation was used to generate the initial 2D contour or not. The 2D slices on which at least one contour was interpolated or manually created, remained for the further analysis. All contours were resampled to increase the point density – the maximum distance between two neighboring points of a resampled contour was 0.01 mm as we wanted to achieve a 0.1 mm measurement accuracy.

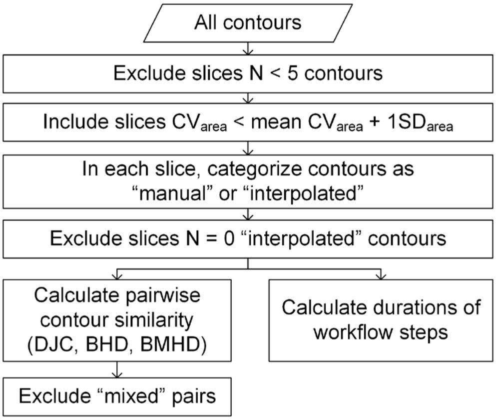


Figure 6-2 The data inclusion and analysis process

Using three different measures, the similarity of contours was evaluated pairwise by a program developed based on the MevisLab® (Kuijf 2015). The Dice–Jaccard coefficient (DJC) (Fotina et al. 2012) was introduced as a measure of the overlap of the enclosed areas between two contours where 1 indicates complete overlap and 0 indicates no overlap. The Bidirectional Hausdorff Distance (BHD) (Huttenlocher et al. 1993) was used to measure the largest variation between shapes of two contours. BHD is defined based on Direct Hausdorff Distance (DHD). Given two contours $C1$ and $C2$, DHD delivers the distance from $C1$ to $C2$ and it can be defined as $DHD(C1, \mathrm{C}2) = \sup_{r\in \mathrm{C}1} (\inf_{s\in \mathrm{C}2} |r - s|)$. In a generalized discreet form, contours $C1$ and $C2$ are available as the point sets P_{C1} and P_{C2}, where $P_{C1} = \{P_{C1}^{i} \in C1 \mid i = 1, m\}$ and $P_{C2} = \{P_{C2}^{i} \in C2 \mid i = 1, n\}$, representing contour $C1$ and $C2$, respectively. Thus the DHD from P_{C1} to P_{C2} is $DHD(C1, C2) = \max_{i=1,m} \min_{j=1,n} \left|P_{C1}^{i} - P_{C2}^{j}\right|$. Though DHD is able find the largest shape variation from contour $C1$ to $C2$, it is directional, i.e., $DHD(C1, \mathrm{C}2)$ is not always same as $DHD(C2, \mathrm{C}1)$. Therefore, we introduced BHD which is defined as $BHD(C1, C2) = (DHD(C1, C2) + DHD(C2, C1))/2$. Similar to the concept of BHD, to measure the average deviation between contour $C1$ to $C2$, we introduced Bidirectional Mean Hausdorff Distance (BMHD), which is defined as $BMHD(P_M, P_E) = \frac{1}{2}(\frac{1}{m}\sum_{i=1}^{m} \min_{j=1,n} \left|P_{C1}^{i} - P_{C2}^{j}\right| + \frac{1}{n}\sum_{i=1}^{n} \min_{j=1,m} \left|P_{C2}^{i} - P_{C1}^{j}\right|)$, as the overall shape similarity measure (Song et al. 2017). $BMHD$ is non-directional regarding contours and comparing to BHD, it is able to reduce the sensitivity to noise and represents the overall shape similarity between contours $C1$ and $C2$.

Measures of the contours were calculated for $C_5^2 = 10$ pairs of contours in each slice. Those pairwise measures were then categorized as being *manual* (both physicians contoured manually), *mixed* (one physician contoured manually, the other used interpolation), or *interpolated* (both physicians used the interpolation tool). The *mixed* pairs were not further analyzed. Independent samples t-test was conducted to evaluate the significance of variation in the mean values using SPSS (version 22).

The details of the software interactions within each slice were extracted from the interaction log files. Each interaction was categorized according to the moment it happened within the steps of the workflow: familiarization, initial contouring, immediate correction, evaluation, and additional corrections. The duration of each of the contouring workflow step was calculated as a sum of the durations of the interactions occurring within this step. Since not all physicians had interactions within each of the five workflow steps, the overall occurrence rate was calculated as a percentage of the total number of engagements in the step over the total number of contouring workflows of the given case. Independent samples t-test was conducted to evaluate the significance of variation among the durations of the workflow steps in the manual and interpolated contour using SPSS.

6.1.4 Results

The subjective ratings of the cases were given by a physician prior to the study as: case C-1 was identified as easy, case C-2 as difficult, and case C-3 as medium difficulty (see example Figure 6-3).

The calculated NASA-TLX indexes corresponded to the rated difficulty levels, though gaps among them were small: the individual NASA-TLX index values being 5.6 out of 20 in case C-1, 7.8 out of 20 in case C-2, and 6.7 out of 20 in case C-3 (Figure 6-4**)**.

In total, 377 GTV contours on 83 slices were created by five physicians over the three cases. Fifteen slices had less than five contours on them and were excluded from further analysis. On the remaining slices, the mean enclosed area of contours in a slice was 448 mm^2 (SD = 199 mm^2) in case C-1, 876 mm^2 (SD = 323 mm^2) in case C-2, and 596 mm^2 (SD = 269 mm^2) in case C-3. In boundary slices towards the superior and inferior directions, the mean enclosed areas Mean$_{area}$ were decreasing as expected.

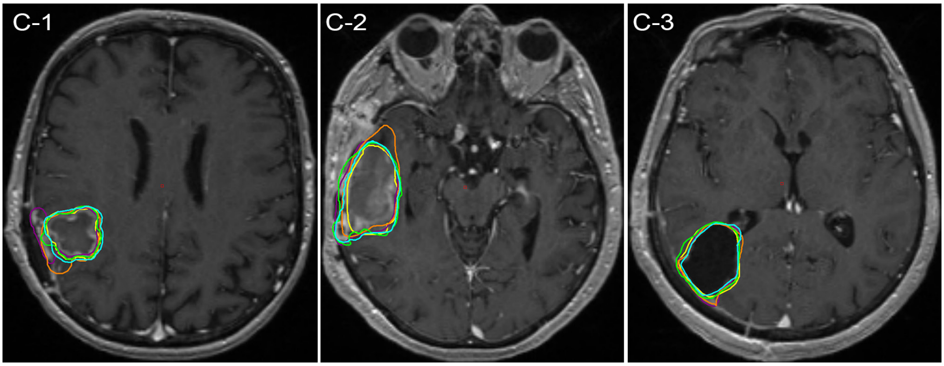


Figure 6-3 Examples of the three cases and the resulting contours on 2D axial slices. The contours of five physicians (each in different color) are overlaid on MRI T1-weighted contrast enhanced image of case C-1, C-2 and C-3

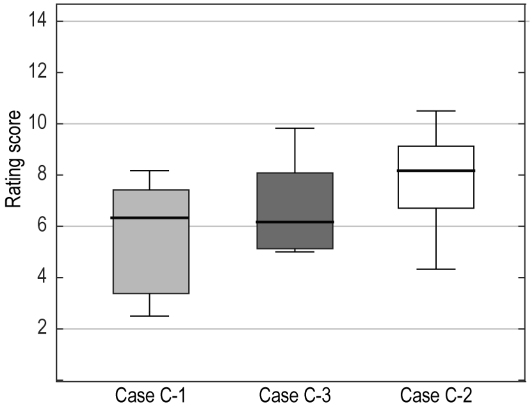


Figure 6-4 The boxplot of the results of NASA-TLX regarding case C-1, C-2 and C-3, the sequence is adjusted according to the mean difficulty levels

The CV_{area} gives a comparable measure of variation of the contoured areas on each slice, with a value of 0 indicating no variation. The mean CV_{area} was 0.18 (SD = 0.17), 0.22 (SD = 0.25) and 0.15 (SD = 0.20) for the three cases C-1, C-2, and C-3, respectively. Based on the CV_{area} and the standard deviation of it, six slices were categorized as “outliers” and were excluded from further analysis. In addition, six slices were eliminated as only containing one type of contours (all manual). As a result, for the detailed analysis, contours on 56 slices remained: 8 slices in case C-1, 23 slices in case C-2, and 25 slices in case C-3, involving 280 individual contours (40 in C-1, 115 in C-2, and 125 in C-3). Among these 280 contours, 144 contours were initiated manually (manual group), and 136 were initiated using the interpolation tool (interpolation group).

6.1.4.1 Influence of automation to the contours

An overview of the calculated measures of the included contours is presented in Table 1.

Table 6-1 Overview of contour similarity measures. The following measures were calculated pairwise between two physicians in a slice within the group: DJC – Dice-Jaccard coefficient; BHD – Bidirectional Hausdorff Distance; BMHD – Bidirectional Mean Hausdorff distance. Mean over these individual pairwise measures is presented in this table. P-values are from the independent samples t-test conducted between these two groups.

Measure	**Grouping**	**Case C-1**	**Case C-2**	**Case C-3**
Subjective Rating		Easy	Difficult	Medium
Number of contours	Manual	26	57	61
	Interpolated	14	58	64
Mean area	Manual	517 mm2	947 mm2	617 mm2
	Interpolated	541 mm2	908 mm2	695 mm2
	P-value	0.607	0.489	0.060
Mean DJC	Manual	0.78	0.72	0.80
	Interpolated	0.87	0.76	0.85
	P-value	0.002	0.155	0.011
Mean BHD	Manual	4.5 mm	10.5 mm	4.2 mm
	Interpolated	2.5 mm	7.8 mm	3.3 mm
	P-value	0.003	0.005	0.005
Mean BMHD	Manual	1.4 mm	2.3 mm	1.3 mm
	Interpolated	0.9 mm	2.1 mm	1.1 mm
	P-value	0.106	0.209	0.038

The overlap between physicians' contours was generally high, with the overall mean DJC being 0.79 (min = 0.30, max = 0.94). In the interpolation group, the overall mean DJC was 0.81, thus being slightly higher than in the manual group where it was 0.77. In the studied cases, the DJC showed a tendency to be on average higher by 0.04 - 0.09 when the interpolation tool had been used. In two of the three cases, the improvement also reached statistical significance (p=0.002, p=0.011 for cases C-1, and C-3 respectively). Such an increase indicated that contours initiated by the interpolation tool were more similar to each other within a slice.

The BHD on a slice was significantly smaller when interpolation was used for all three cases (p = 0.003, p = 0.005, and p = 0.005). The decrease was the highest in case C-2, where it was reduced by 2.7 mm, followed by C-1 where it was less by 1.5 mm and the smallest reduction was in case C-3 by 0.9 mm. In terms of the overall shape similarity as measured by BMHD, the average distance between the two contours, independent from its creation method, was 1.2 mm in the cases C-1 and C-3, and 2.2mm in case C-2. Generally, the mean BMHD showed a tendency towards a decrease when the interpolation had been used but was only significant in case C-3 (p = 0.038).

6.1.4.2 Influence of automation on the contouring process

Detailed contouring workflow within a slice as was observed in the conducted study is depicted in Figure 6-5. The initial contouring step (Step 2A or 2B) represented the action of creating the first (i.e., initial) closed loop boundary of the visible tumor, visually inspecting and perceiving the contour and/or the medical image(s) while contouring, as well as of deciding on the next action (i.e., to correct the contour or to navigate away). Contour corrections were categorized as immediate corrections and additional corrections. The immediate corrections (Step 3) accounted for the corrections of the contour until the first slice change (i.e., navigate away). These corrections were done, for example, to compensate for mouse inaccuracy (Zabramski 2011), or to adjust the contour based on the further inspection of the presented 2D medical image(s) as well as clinical reasoning. Returning to the contour for corrections after inspecting the neighbor slices or at any later moment, were identified as additional corrections (Step 5).

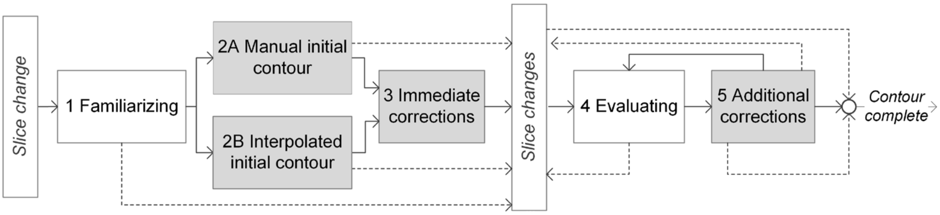


Figure 6-5 The contouring workflows of this study with a variation in the initial contour creation step. In the manual workflow, as step 2A the physician manually contoured the boundary of the tumor. In the interpolated workflow, as step 2B the physician used the between contour interpolation tool. Data regarding the contouring process was extracted according to these workflow steps

The mean durations of the workflow steps are presented in Figure 6-6. Generally, a physician completed the contouring task faster when using the interpolation tool. In terms of specific workflow steps, when interpolation was used, physicians tended to spend more time on familiarizing (Step 1) and less time on evaluating (Step 4). Furthermore, some physicians tended to spend more time to complete the task compared to others.

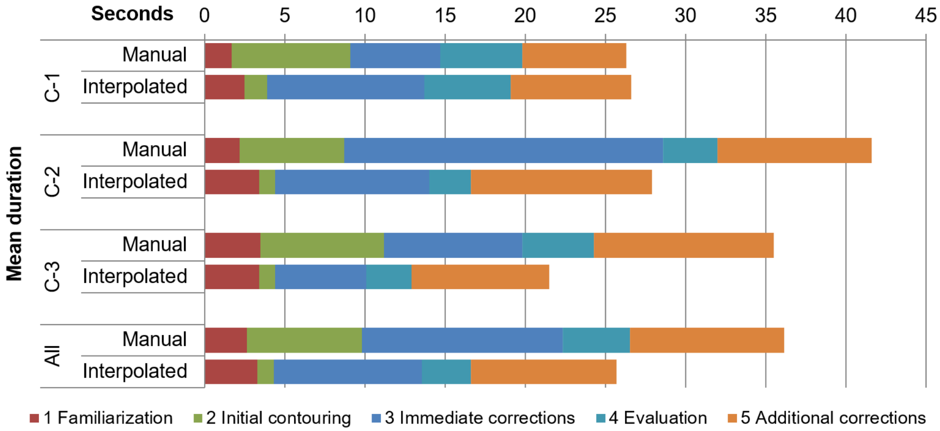


Figure 6-6 The mean durations of different workflow steps in case C-1, C-2, and C-3. The type of workflow is labeled as manual or interpolated. In addition, the average of each step over all cases is shown as "All"

The details of the workflow steps averaged over all physicians for each case are shown in Table 6-2**.** In addition, within each workflow, the total durations of contour corrections (sum of time spent on Step 3 and Step 5) were calculated. The total durations of the contouring process on a slice per physician were also summed. Furthermore, the average duration of each step was also calculated over the three cases.

Table 6-2 Details of the workflow steps of the three cases. Occurrence percentage, the mean duration (in seconds per slice), and standard deviation (SD) of the duration of each workflow step for both types of workflows are listed. Step 2A and 2B were not representing the same type of interactions, thus statistical comparison was not suitable (N/A). s = second

Workflow		C-1		C-2		C-3		All	
Step	Type	Occurrence	Mean duration (± 1 SD)	Occurrence	Mean duration (± 1 SD)	Occurrence	Mean duration (± 1 SD)	Occurrence	Mean duration (± 1 SD)
1 Familiarizing	Manual	100%	1.7 s (± 2.7 s)	100%	2.2 s (± 1.4 s)	100%	3.5 s (± 5.2 s)	100%	2.7 s (± 3.7 s)
	Interpolated	100%	2.5 s (± 1.3 s)	100%	3.4 s (± 2.1 s)	100%	3.4 s (± 3.3 s)	100%	3.3 s (± 2.7 s)
	p-value	-	0.303	-	0.001	-	0.937	-	0.100
2A/B	Manual	100%	7.4 s	100%	6.5 s	100%	7.7 s	100%	7.2 s

Initial contour			(± 2.2 s)		(± 1.9 s)		(± 3.3 s)		(± 2.7 s)
	Interpolated	100%	1.4 s (± 0.9 s)	100%	1.0 s (± 0.8 s)	100%	1.0 s (± 1.3 s)	100%	1.0 s (± 1.0 s)
	p-value	-	N/A	-	N/A	-	N/A	-	N/A
3 Immediate corrections	Manual	8%	5.6 s (± 3.8 s)	7%	19.9 s (± 14.0 s)	7%	8.6 s (± 6.5 s)	7%	12.5 s (± 11.0 s)
	Interpolated	50%	9.8 s (± 3.1 s)	28%	9.6 s (± 9.5 s)	5%	5.7 s (± 3.2 s)	19%	9.2 s (± 7.7 s)
	p-value	-	0.142	-	0.095	-	0.468	-	0.316
4 Evaluating	Manual	100%	5.1 s (± 4.0 s)	100%	3.4 s (± 2.5 s)	100%	4.5 s (± 3.3 s)	100%	4.2 s (± 3.2 s)
	Interpolated	100%	5.4 s (± 7.0 s)	78%	2.6 s (± 4.8 s)	89%	2.8 s (± 3.2 s)	85%	3.0 s (± 4.5 s)
	p-value	-	0.853	-	0.347	-	0.004	-	0.018
5 Additional corrections	Manual	46%	6.5 s (± 7.0 s)	46%	9.6 s (± 7.5 s)	39%	11.2 s (± 9.4 s)	43%	9.6 s (± 8.2 s)
	Interpolated	57%	7.5 s (± 5.3 s)	22%	11.3 s (± 9.8 s)	45%	8.6 s (± 6.6 s)	37%	9.1 s (± 7.4 s)
	p-value	-	0.753	-	0.567	-	0.242	-	0.723
(3+5) Total correction	Manual	50%	6.9 s (± 7.0 s)	51%	11.4 s (± 10.4 s)	43%	11.7 s (± 9.2 s)	47%	10.6 s (± 9.4 s)
	Interpolated	79%	11.7 s (± 6.1 s)	40%	13.1 s (± 14.0 s)	48%	8.6 s (± 6.3 s)	48%	10.7 s (± 9.8 s)
	p-value	-	0.092	-	0.623	-	0.144	-	0.969
Total	Manual	-	17.7 s (± 9.9 s)	-	18.1 s (± 12.0 s)	-	20.8 s (± 13.5 s)	-	19.2 (± 12.3 s)
	Interpolated	-	18.6 s (± 14.3 s)	-	12.0 s (± 15.9 s)	-	11.0 s (± 8.9 s)	-	12.2 (± 13.0 s)
	p-value	-	0.818	-	0.022	-	<0.001	-	<0.001

When the initial contour was done manually, physicians always returned to the slice (i.e., 100% occurrence of the evaluation step). No separate evaluation step was recorded in 15% (N = 20) of all contours initiated by the interpolation tool, which indicated that after the initial contour was interpolated, and possibly corrected (N = 2), the physician did not revisit it. Regarding individual cases, such contouring workflows were present in two of the cases, being 22% (N = 13) in case C-2 and 10% (N = 7) in case C-3. Further analysis revealed that the average viewing time of those interpolated contours was 0.6 s (SD = 0.19 s), which is less than the overall average of 1.0 s. More than half of such contours (N = 9 in C-2, N = 4 in C-3) could be accounted for one physician.

6.1.5 Discussion

6.1.5.1 Automation bias

Automation may influence physicians' reasoning during contouring by providing an automatically generated contour. When such a contour is accepted without sufficient evaluation of the available data, automation bias occurs and errors might be introduced. Automation bias may have either negative or positive effect on the process and the outcomes of the contouring task, as in many steps of the contouring task, physicians must make a subjective decision based on their knowledge and experience.

The influences of automation on the reasoning process are more difficult to be categorized as being positive or negative. One of the challenges in evaluating the outcomes of a contouring task is that there is no gold standard in GTV contouring (Weiss and Hess 2003). There is general acknowledgement that less variation among physicians is desired, i.e., methods which lead to reduced inter-observer variation with improved consistency are preferred. However, categorizing variations to be erroneous is challenging due to the nature of task. Another aspect that can be measured is the amount of time spent on inspecting data as shown in the *Familiarizing* and *Evaluating* steps of the workflow. However, increased time does not necessarily correlate with the quality of contours as physicians are capable of detecting abnormalities rather rapidly (Drew et al. 2013).

In this chapter, inter-observer variation of selected contours was used to evaluate effects of automation bias on the outcomes. On the negative side, the automation bias may lead to errors in the contours. On the positive side, it may increase consistency of contours. The inter-observer variation can be evaluated by different types of measures such as the DJC (area overlap), BHD (shape outliers) and BMHD (shape similarity), where smaller variation among physicians indicates higher confidence in having the "consistent" tumor contour. Regarding the process, the mean durations of different steps of the task were adopted as the measures of effects of automation bias.

6.1.5.2 Inter-observer variation among outcomes

In radiotherapy, 60% or more of the mis-administrations are due to human error (Duffey and Saull 2002). Lack of a gold standard, as well as the expected variation among physicians, increases the probability of human errors during the contouring task. For instance, Brundage et al. (Brundage et al. 1999) identified that insufficient target volumes were one of the common reasons for treatment plan modification. In order to tackle this, in clinical practice, peer review is the proposed approach to decrease the probability of such (and other) human errors (Marks et al. 2013; Mackenzie et al. 2016; Brunskill et al. 2017). In short, it is expected that the smaller the variations among physicians are, the fewer errors there are.

Variation among physicians is well documented (e.g., (van Herk 2004; Louie et al. 2010; Fotina et al. 2012; Dinkel et al. 2013; Whitfield et al. 2013)). However, there is a lack of consensus on which measures to use for judging the variability (Fotina et al. 2012). Furthermore, there is no reproducible gold standard for evaluating the accuracy of contours due to many reasons (e.g., image quality, and subjectivity of physicians (Weiss and Hess 2003)). In many studies, a manual contour done by an experienced physician (i.e., expert contour) is being used as a reference (Olabarriaga and Smeulders 2001). Such an approach can be sufficient to evaluate the reproducibility of an automatic contouring method, but the results are dependent on the contours provided by that expert. This study aimed to measure whether the manually initiated contours were more similar to each other than contours initiated by the interpolation tool. Thus, we incorporated pairwise contour similarity measures such as pairwise DJC, pairwise BHD, and pairwise BMHD. Based on the results presented in the results section, we observed a tendency that contours initiated by the interpolation tool were slightly more similar to each other among different physicians than manually initiated contours. In all three cases, the mean BHD and BMHD decreased, while also the mean DJC showed improvement. Statistical significance was reached for six of the nine pairwise calculated similarity measures. One of the sources for the increase is shape similarity might be that the computer is better in creating a smoother shape compared to the human, who must draw it manually with a mouse in this study.

Though the shape similarities of the GTV contours were improved, the improvements were below the current accuracy of radiotherapy. For instance, we observed a mean shape variation (measured by BMHD) decrease by 0.2 mm – 0.5 mm. In the treatment plan of GBM, the recommended margin to encompass possible treatment delivery uncertainties is between 3-5 mm, depending on the specific situation (Niyazi et al. 2016). Such margins are used to compensate the uncertainties in the GTV contouring as well as for shifts in patient positioning. For instance, Drabik et al. (Drabik et al. 2007) measured that on average there was an (up to) 0.5 mm positioning shift of a GBM patient in the treatment. Nevertheless, among multiple sources of uncertainty within the radiotherapy planning process (van Herk 2004), GTV contouring has been identified as the weakest link (Njeh 2008). Thus, decreasing variation in GTV contouring can be beneficial especially that the level of precision of dose delivery is increasing (Schaffner and Pedroni 1998).

The case difficulty could not be clearly associated with reduced variations of the contours initiated by the interpolation tool. The simplest case (C-1) showed the largest improvement, while the medium difficulty case (C-3) and difficult case (C-2) showed similar tendencies. Therefore, further studies with more cases of varying levels of difficulty are required to evaluate the correlation between the decrease of variation by utilizing automation and the difficulty of the case. At the same time, it was clear that the level of difficulty is related to the general level of variation among physicians. The more difficult case in the study (C-2) had the lowest DJC and the highest BHD. Besides, the BMHD in this case was nearly double compared to the other two cases.

6.1.5.3 The efficiency of and the influences on the contouring process

Detailed analysis of the contouring process reveals the impact of incorporating automatic initial contour creation (i.e., interpolation) to the overall process. The between slice interpolation tool that was investigated in this study, changed the way the initial contour was created (click of a button or press of a key on keyboard, instead of drawing with the mouse). As expected, including automation generally decreased the overall contouring time. In the case C-1, the average duration was slightly higher, though not statistically significant. For this specific case, it might have been influenced by the small size of the tumor, larger slice thickness (2.5 mm instead of 1.25 mm), or being an easy case. In the case C-2, the overall duration was reduced due to the shorter initial contour drawing time. In the case C-3, the evaluation step was also significantly shorter when the interpolation tool had been used, resulting in a further reduction of the task completion time.

The availability of the interpolation tool, for some physicians changed their contouring strategy. During this interpolation-influenced contouring strategy, the physician would first contour in a set of slices manually while skipping some in-between slices (i.e., seeing them but not contouring on them), and then return to the empty slices later in the process and utilize interpolation to fill in the missing contours. This type of contouring strategy is characterized by slight changes of the contouring workflow on the interpolated slices: longer time may be spent in familiarizing (step 1), fewer additional corrections (steps 3 and 5) on the interpolated contours, and there are fewer (or no) returns (step 4) to the slice once interpolation had been used.

The frequency of corrections gives a measure of the acceptance of the contour. Based on the presented three cases, it was observed that the frequency of corrections (on average 47.5% of the cases), as well as the duration of them, remained similar for both manually drawn initial contours (47%) and contours initiated by the interpolation tool (48%). This could indicate that if a contour is in a clinically acceptable range, then the likelihood of a manual correction is independent from its' original creation method. Eighty-four percent of these corrections occurred after returning to the contoured slice at a later point. One common motivation for correction, for example, is a comparison with neighboring slices (Aselmaa et al. 2017). These later stage corrections can be assumed to correspond to the physicians updating their mental model (Varga et al., 2013) and then correcting the contours correspondingly.

In medical image related decision-making, the duration of one second is considered to be a significant allocation of visual attention for detecting an object of interest (Hillstrom 2000). In addition, it has been shown that a visual fixation time of one second is significantly correlated with correct detection of a lesion (Nodine et al. 2002). In our study, an interpolated contour was on average viewed for one second prior to an action, indicating that the level of evaluation for determining the correctness of a contour could be deemed sufficient. In the study, 15% of the interpolated contours were not revisited. In the contouring process of those contours, the physician spent on average 0.6 s viewing it prior to changing to another slice, being below the recognized sufficient level of visual attention allocation. However, this measure on its own is not sufficient for concluding whether this 0.6 s is a sufficient duration of visual inspection in such specific cases. At the same time, interpolated contours showed a slight improvement in the inter-observer variation. Thus, even though the automation bias seems to be present, it was leading towards more desirable results and reductions in the overall task completion times. Therefore, the use of interpolation can be encouraged.

6.1.5.4 Pros and cons of automation

The reasoning occurring during the contouring task is influenced by a number of variables, such as the type of treatment, whether there was a preceding surgery, the size and the location of the tumor, tumor characteristics, etc. (Aselmaa et al. 2014). Physicians need to weigh such various aspects against their past experiences in order to reach a decision. This process can be seen as case-based reasoning where individual knowledge captured from a very specific context (e.g., treating a particular patient with a particular disease) can be extrapolated to similar contexts (Pantazi et al. 2004).

The benefit of a (semi-)automatic contouring method strongly depends on its robustness. For example, during this study, in few instances, the interpolation generated a partially zig-zag contour instead of a smooth one which took physicians' more than average efforts to correct. Automatically generated contours, that are found unacceptable, result in unnecessary software interactions and thus could increase workplace frustrations. It has been reported that in general there is a rather high loss of work time due to frustrating experiences with software (Lazar et al. 2006) which in turn led to higher financial costs and possibly even impacts the outcomes of the treatment (Johnson 2006). Therefore, advances in improving the robustness and increasing the accuracy of contouring methods, together with improving the general usability of software solutions are required.

In our study, it was identified that automation guided physicians towards more similar contours, which is a desired effect as there is no gold standard. We postulate that when the automation is used to provide contouring aids on 2D slices, the automation bias is more noticeable on the slices where the level of cognitive involvement is lower. At the same time, automation bias can be more prominent in the more cognitively demanding situation, but may be obfuscated by other variables influencing the physician's subjective reasoning process.

6.1.5.5 Limitations

The study presented was conducted on three different patient datasets. Conducting a study involving manual contouring is challenging due to the time requirement from the physicians. However, a larger sample size would be beneficial to have a deeper understanding the influence of automation bias in relation to other variables such as the size of the tumor, slice thickness, levels of case difficulties, or levels of physician's experience.

The case difficulties were based on a subjective rating of a senior physician acquired independently from the present study. Those ratings were given in three-point scale (easy, medium, difficult). A more robust evaluation method for determining case difficulty could be beneficial, for example, objective description of the tumor based on image features (Gevaert et al. 2014).

The aim of this study was to investigate the automation bias in a naturalistic setting. While we found our findings valuable, a controlled study with fewer variables (e.g., pre-defined choice of the tool per physician) may reach stronger conclusions. In addition, though this study describes the relations between automation bias and the reasoning process based on the software interaction data, studies complemented with eye tracking might reveal more insights of the influence of automation on the reasoning process.

6.1.6 Conclusion

Automation is increasingly incorporated into the radiotherapy planning process. This chapter presented a study of evaluating the impact of using a between slice interpolation for initiating a contour on the resulting contours as well as on the contouring process in comparison to the fully manual contouring.

A GTV contouring study with five physicians on three patient cases was conducted, from which 280 individual 2D contours were analyzed. The contours obtained with and without the use of the interpolation tool were pairwise analyzed within each slice in terms of area overlap (DJC), shape outliers (BHD), and overall shape similarity (BMHD). In all measures, outcomes based on the use of the interpolation tool showed an increased agreement among physicians (DJC increase by 0.04 – 0.09; BHD decrease by 0.9 mm – 2.7 mm; BMHD decrease by 0.2 mm – 0.5 mm).

Influences to the contouring process were also identified. The efficiency was improved – the overall interaction time within a slice was reduced by 6.1 seconds ($p = 0.022$) and 9.8 second ($p < 0.001$) in two of the three cases, mainly due to the time-saving in creating the initial contour. In addition, interpolated contours were corrected at a similar rate as manually drawn contours, which indicated a similar level of evaluation. In a sub-set of contouring processes, an interpolation influenced contouring strategy was identified. This contouring strategy consisted of first contouring in a set of slices manually and then used the interpolation tool to fill in the missing contours in the in-between slices. However, precaution is needed, as in our study 15% of interpolated contours were not revisited after initial creation and inspection.

Based on the presented findings, it can be concluded that using the between slice interpolation tool influences the contouring outcomes in a desirable direction, as well as significantly decreases task completion time. Thus, the use of such automatic contouring tools can be encouraged in radiotherapy planning software.

Chapter 7

Discussion and conclusion

This chapter reflects on the main outcomes of the research presented in this thesis and discusses their relevance. We also offer suggestions for future research directions.

The research presented in this thesis evolved through our exploration of the radiotherapy context from both cognitive and interaction design perspectives. We conducted research ranging from general radiotherapy workflow to treatment planning (Chapter 2), to the contouring task (Chapters 3 and4), to the tumor (i.e. gross tumor volume, GTV) contouring task (Chapters 5 and 6). We reveal insights into physicians' sensemaking and general cognitive processes throughout this discussion.

We employed a number of research methods, including ethnographic studies, interviews, literature review, and prototype-based studies. We then propose a two-step approach for incorporating sensemaking theory into the design process for support software (Chapter 5). We demonstrate that this approach is useful for gaining a detailed understanding of physicians' sensemaking processes during GTV contouring and for identifying requirements for sensemaking support software. Furthermore, we investigated the influence of automation (Chapter 6) and found that software design influences physicians' reasoning process as well as the resulting contours.

7.1 Reflections

7.1.1 Research approach

Our exploratory research commenced with observational studies in hospitals and was complemented with interview-based studies. In this phase of research, our aim was to gain an understanding of the context and define our research focus. The next stage involved conducting studies with physicians. In a technology driven context such as radiotherapy, however, conducting user studies often requires highly functional prototypes (Sauer, Seibel & Rüttinger, 2010). This requires the design researcher to have appropriate programming skills or to collaborate closely with a researcher, software engineer, or software engineer-researcher. While it is reasonable to expect researchers to be able to program to some degree, there are also indications that designers are becoming more familiar with programming (Dorn & Guzdial, 2006) and that design researchers too will increasingly need to have programing skills.

One of the challenges when researching physicians' sensemaking process is to develop a systematic methodology. Dou et al. (2009) demonstrated that it is possible to a large extent to identify different reasoning strategies based on software interactions. Therefore, the approach we adopted on this research project was to record physicians' interactions with the software prototype while performing the contouring tasks. We then visualized the interaction data collected as individual timelines representing the task process. We subsequently used the visualized timelines together with the interaction details to make certain assumptions regarding the sensemaking process (Chapter 5). Interestingly, Kannampallil, Abraham & Patel (2016) recently proposed a methodological framework for a process-based approach to capture, analyze and gain insight regarding the processes of interest. Our methodological approach was in line with and supported by this proposed framework.

7.1.2 Radiotherapy

7.1.2.1 Workflow analysis

We started with an analysis of the complete radiotherapy workflow (Chapter 2). This served two purposes: first, to define the scope of our research by identifying problem areas that could benefit from design research. The workflow analysis enabled us to gain a deep understanding of the challenges of the contouring task, and how this task relates to the other steps in the radiotherapy workflow.

The second purpose of the workflow analysis was to establish an area of common ground within the research consortium. The workflow we put together served as a communication and co-design tool (Freudenthal et al., 2011). It also helped to position the individual research projects of ten young researchers and four experienced researchers and to identify possible areas for collaboration. We therefore highly recommend conducting workflow analysis in the earliest phase of a collaborative research project.

7.1.2.2 Contouring strategies

Mapping different contouring strategies was not the primary aim of this research project; however, we did observe some such strategies while analyzing contouring processes. For instance, we noted that some physicians are more systematic in their approach than others (Chapter 5). One of the contributing factors to the choice of contouring strategy may be the personality traits of individual physicians (Saposnik et al., 2016). In addition, we identified that, on some occasions, the availability of an automatic 2D contouring method also influenced the overall contouring approach (Chapter 6). However, further investigations are needed to identify possible correlations with sensemaking approaches, contouring methods used, the extent of contour modifications, and physicians' level of experience.

7.1.3 Sensemaking theories

Our research focused on the cognitive process of 'sensemaking' – a term that has been used in both a narrow and a broad sense (Zhang & Soergel, 2014). Sensemaking in the narrow sense is about building cognitive frames, interpretation, and comprehension. In the broad sense, it also incorporates the activities of information gathering, searching and filtering etc. (i.e. information foraging). These two uses of the same term can be confusing; however, both are valid and applicable in different situations.

In Chapter 3, we presented a model of external aspects of individual sensemaking in the narrow sense. This model highlighted the relationships between various cognitive processes such as perception, information foraging, problem-solving and decision-making. Immersing ourselves deeper into the topic we realized that, from an interaction design perspective, it is difficult to make a separation between software interactions that support sensemaking (in the narrow sense) and information foraging. We therefore adopted the broader definition of sensemaking (Chapter 5), which afforded us a more holistic view of the process.

In the context of designing medical imaging software, data presentation is of great importance. This topic is being extensively researched in the field of information visualization because it relates to representing data so that it is in line with human cognitive processes (Patterson et al., 2014). Sensemaking in the narrow sense may be more useful in such a field.

7.1.4 Interaction design

The need for software to be highly usable is well established. For instance, a search in the Scopus database (Scopus, 2017) for 'usability' or 'usable' in the title, keywords, and abstract turns up more than 91,000 publications (e.g. journal articles or conference publications) and 1.58 million patents. It has been shown that high usability of software increases user satisfaction and is consequently financially beneficial for the software vendor (Karat, 1990 and Donahue, 2001). There is an abundance of methods and guidelines available for designers to apply to ensure their software products are usable (e.g. Nielsen, 1995 and Kushniruk & Patel, 2004). The human-centered design approach is the prominent direction for ensuring usability and is used in the development of international standards (ISO 9241-210: 2010 and IEC 62366-

1: 2015). The increasing amount of data relating to performing clinical tasks poses new challenges for designers. Focusing on usability is no longer sufficient, designers also need to comprehend and support clinicians' cognitive processes.

Designers increasingly need to understand and incorporate technological advancements (e.g. automation); however, the way in which software is designed may intentionally or unintentionally influence physicians' performance (Rosman et al., 2014). Designers need to be aware of this influence and avoid elements that lead to negative consequences, and embrace those that lead to positive outcomes. In Section 5.1, we describe our exploratory study of software design in line with the sub-region based contouring workflow (Section 5.1). However, it must be noted that the sub-region based contouring workflow leads to smaller contours and is generally treated with caution by physicians. As part of this study, we also evaluated the impact of automatic contouring (Chapter 6). In general, we saw improved efficiency when completing the contouring task as well as greater consistency, which is another desired effect.

One of the intended outcomes of this research project was to identify requirements for sensemaking support software. We employed different research methods which enabled us to identify several requirements and design considerations. An overview of these requirements is provided on pp. 216-218. It is worth mentioning that we have used the term 'requirements' rather broadly – they should really be viewed as high-level requirements more than actionable requirements. Actionable requirements should be justified, clear, unambiguous and verifiable (Mannion & Keepence, 1995).

7.2 Recommendations for future research

7.2.1 Research methods

The research presented in this thesis was based on studies conducted with a limited number of physicians (five to eight) and tasks (three cases). Stronger conclusions could be reached with a higher number of participants or tasks.

The contouring step in the radiotherapy workflow generally involves multiple physicians; typically, a resident (physician in training) does the majority of the contouring and the senior physician revises, corrects and validates those contours. Informal discussions (Munoz et al., 2011) and/or planning review discussions also take place (Marks et al., 2013). Our study did not include such revision processes even though peer review has been shown to improve the quality of contours (Mackenzie, Graham & Olivotto, 2016). Research into the collaborative sensemaking process when using software could certainly offer some interesting insights. Collaborative sensemaking has otherwise been examined in contexts such as hospital emergency departments (Paul, Reddy & DeFlitch, 2008), asynchronous visual analytics (Chen et al., 2011), and information searching tasks (Bhavnani, Clarkson & Scholl, 2008).

We visualized user interaction data captured during our research and on this basis identified interaction patterns. Human pattern recognition is an advanced approach (Mattson, 2014) and as such is feasible for use with small sample sizes. For larger or more complex studies incorporating machine learning methods, we would recommend identifying behavioral patterns or even inferring sensemaking activities (Kodagoda et al., 2016). In addition, richer insights into sensemaking could be uncovered utilizing eye-tracking (Al-Moteri et al., 2017), posture (DiDomenico & Nussbaum, 2005), electroencephalography (EEG) (Chaouachi, Jraidi & Frasson, 2011)), heart rate variability (Galy, Cariou & Melan, 2012), galvanic skin response (Shi et al., 2007), and/or verbal protocols (e.g. think-aloud (Van den Haak, De Jong & Schellens 2003)).

7.2.2 Contouring in radiotherapy

The contouring step in the clinical workflow consists of contouring OARs, GTV, CTV, and PTV (Chapter 2). There is a further level of complexity to contouring all these volumes that we did not investigate in detail in this project, namely that [editor’s note: please specify]. Furthermore, in the advanced treatment plans, there may be multiple GTV volumes that represent metabolic abnormalities, for example (Ken et al., 2013). Further research is needed to identify the types of sensemaking support required for a combination of dependent and intertwined tasks of this nature.

Contouring the CTV can be cognitively challenging since this volume is intended to represent the spread of the tumor that is known but ‘invisible’ on the medical images. In a typical scenario, CTV is created automatically by applying a 3D expansion of the GTV by a margin and then is corrected as needed. For example, in the case of GBM, this margin can be 2-3 cm (Mason et al., 2007). The cognitive difficulties lie in judging this generated CTV contour. There is a growing body of clinical knowledge that aims to find new ways of overcoming these uncertainties. For example, new types of medical images may provide relevant information to assist in this regard (e.g. functional MRI

(Drevelegas & Papanikolaou, 2011)). These clinical findings may also have implications for the design of radiotherapy software solutions.

Contouring of OARs is considered less cognitively challenging than tumor contouring, primarily because there is less variation between patients, the organs are expected to be a certain size and shape, and anatomical knowledge can be applied during the contouring. For this reason, most research on OAR contouring has focused on improving efficiency and decreasing the effort required from physicians. However, there can still be cognitive difficulties in detecting some regions of OARs which also need to be addressed (Ramkumar et al., 2016).

7.2.3 Automation in radiotherapy

The effect of automation on physicians' sensemaking requires further investigation. For this research project, we conducted a study on one type of basic automatic contouring method (Chapter 6); however, these methods vary in their degree of automation (Heckel et al., 2013). Investigations into each of these methods would make it possible to detect whether and when automation introduces a bias. Furthermore, a higher level of automation in combination with automation of multiple contouring tasks may have a cumulative effect of clinical significance. In addition, a study with a larger sample would be beneficial to model the influence of automation bias in relation to other variables such as the size of the tumor, the thickness of slices, the difficulty level of cases, and physicians' level of experience.

7.2.4 Design research

Our work was limited to the design of the software solution and did not address different input devices (e.g. mouse vs pen, screen vs. touchscreen), which has been shown to influence physicians' performance (Ramkumar, Varga, Laprie et al., 2013; Multi-Institutional Target Delineation in Oncology Group, 2011). Studies have also been carried out that compare the overall mental workload between different input devices (Hart 2006); however, the influence of input devices on cognitive processes, such as sensemaking, is yet to be investigated.

One of the major challenges in interaction design is fulfilling all software requirements without compromising usability. Our research was conducted using a software prototype with limited functionality, which is common approach for design research (Stolterman & Wiberg, 2010). In order to evaluate the potential level of benefit sensemaking support software, extensive studies with software covering the full range of necessary functionalities would be required. However, creating biomedical software is a demanding task and beyond the scope of an individual Ph.D. research project. It would require a study in close collaboration or within a company that creates radiotherapy contouring software.

7.3 Vision for contouring software that supports sensemaking

While we have seen that the contouring task benefits from more efficient contouring tools, one future challenge will be to support physicians' sensemaking and reasoning when using the software as well as integrating automation. This thesis is intended as a first step in understanding the context of use and identifying the high-level requirements for sensemaking support software.

We envision that contouring software in the future will employ a number of automatic contouring methods of varying levels of automation. These methods produce high quality results for OARs, which allows physicians to focus on tumor contouring. The software will incorporate a knowledge base relating to each of the automatic contouring methods, such as applicability criteria, success rates, data requirements, required levels of user involvement, etc. In addition, data regarding corrections to contours generated by the contouring method must be gathered and analyzed (e.g. using machine learning algorithms). Furthermore, in cases where multiple methods are applicable, the software will enable the best portion of each automatically proposed contour to be intelligently merged.

The role of physicians in the contouring step of the radiotherapy workflow will be primarily to review the contours and correct them as needed. The software will guide physicians through the process by providing cues on regions to review as well as appropriate data to support those reviews. Medical literature and contouring guidelines will be available on demand to support evidence-based contouring. Once a physician has completed the contouring task, the final contours will be shared with peers who can easily annotate them and make corrections, which then can be either accepted or rejected.

7.4 Conclusions

The aim of this research project was to provide a means of comprehending physicians' sensemaking process during the early phase of software design and was guided by six research questions.

RQ-1. What is the workflow in radiotherapy?

In Chapter 2, we set out the complete workflow for radiotherapy illustrated by a visual diagram (page 77). This workflow involves a number of tasks and clinicians, many of which are dependent on others. It consists of four main phases: diagnosis, treatment planning, treatment delivery, and post-treatment follow-up. Our analysis of the treatment planning phase was presented in Section 2.2. We identified that improvements are needed from both a technological and a human perspective. The immediate areas for improvement are increasing effectiveness and time efficiency, and supporting both cognitive and physical tasks.

RQ-2. What are the cognitive processes involved in the contouring task? Which medical factors influence contouring?

In Chapter 3, we provided an overview of the main cognitive processes involved in clinical reasoning and identified that sensemaking is the underlying cognitive process during the contouring task. We present the results of a study we conduced to explore medical factors that influence reasoning. We divided these factors into three categories: treatment context, tumor context, and tumorous areas. In total, we identified 29 factors, for instance, the type of radiotherapy treatment, the institute administering the treatment, the type of tumor, the proximity of the tumor to the organs at risk, the anatomic barriers of the tumor, and the size of the tumor.

RQ-3. What are the challenges for incorporating automated contouring into software design?

We identified that automated contouring plays an important part in contouring software now and in the future. Therefore, we conducted a modeling-based analysis to identify implications for software design (Chapter 4). We defined four categories of software requirements: general usability, navigation, compatibility with task workflow, and flexibility of interactions.

RQ-4. How can we incorporate sensemaking theory in the early phase of software design?

We developed a two-step approach for incorporating sensemaking theory into the software design process, which is presented in Chapter 5. The steps in this approach are: (1) model the sensemaking process of physicians while performing a task (i.e. develop a contextual sensemaking model), and then (2) identify software interaction design requirements that support sensemaking based on this model. We then put this approach into practice, applying it to GTV contouring. We identified five main areas for design improvement: (1) to enable the development of effective initial cognitive frames; (2) to support intuitive navigation within and between datasets; (3) to support detection of regions of interest; (4) to enable additional methods for contour evaluation; and (5) to improve general efficiency by reducing time and physical effort.

RQ-5. What sensemaking process do physicians follow during tumor contouring?

In Chapter 5, we described the details of the sensemaking process during tumor contouring. We identified three main phases of the tumor contouring task: familiarization, action, and evaluation. We observed that, during these phases, there were differences in the sensemaking activities, although some software interactions remained the same. In addition, we noted that on average 39 per cent of contouring time was spent on navigation. These navigation interactions supported physicians with building the cognitive frame of the tumor, as well as exploratory and focused information seeking.

RQ-6. How does automated contouring influence physicians' cognitive processes?

We noted that automation is an important aspect of present and future contouring software. However, it remains unclear whether automation influences physicians' cognition. Chapter 6 presents a study in which we evaluated one type of automated contouring to address this question. We demonstrated that automation generally improves efficiency but also changes the sensemaking process and consequently influences the outcomes of the task (i.e. the contours).

7.5 Requirements for contouring software that supports sensemaking

Here, we provide an overview of various high-level requirements for the design of software to support sensemaking. Some of these requirements are generalized based on our findings from analyzing the tumor contouring task.

Table 7.5-1 Overview of software design requirements for supporting sensemaking during the contouring task in relation to the associated task phases

	Contouring task phase		
	Familiarization	Action	Evaluation
Allow customizability	✓	✓	✓
Support rapid data exploration	✓		✓
Support frame building	✓		✓
Integrate medical knowledge	✓	✓	✓
Support identification of regions of interest (ROIs)	✓	✓	✓
Facilitate intuitive navigation	✓	✓	✓
Integrate intelligent Boolean operations with volumes		✓	
Incorporate automatic contouring		✓	
Ensure usability of contouring tools		✓	
Support comparison of contours			✓

Allow customizability: There are significant differences between different medical institutes and the types of treatments they administer (Chapter 3). Software should incorporate the procedures described in the protocols of the relevant institute. The design should also be sufficiently flexible to allow institution-level customization without engendering administrative overheads.

Support rapid data exploration: The starting point of the contouring task is familiarization with the data, for which physicians need to be able to examine multiple datasets in an efficient way. Towards the end of the task, physicians must again navigate rapidly through the data to evaluate the completeness of the contours (Chapter 5).

Support frame building: The way in which data is presented to physicians influences their reasoning. This gives the designer the power to encourage or discourage certain behaviors. For instance, one of the activities during the familiarization phase of contouring is identifying the relevant datasets to examine. A corresponding design consideration could be suggesting default layouts of data on GUI (Chapter 5). Another way to support frame building is by proposing contours automatically based on the available information (Chapter 6).

Integrate medical knowledge: During our observational studies, we noted that the physicians frequently referred to contouring guidelines printed on paper. This information could be incorporated into the software design so that, when in doubt, physicians would have easy access to the guidelines for that specific region (Chapter 2).

Support identification of regions of interest (ROIs): A specific aspect of medical knowledge that should be integrated into the software is which elements should or should not be included in ROIs. It is often difficult for algorithms to indicate with 100 per cent certainty which regions should be included or excluded. As an alternative, a guiding system could be incorporated into the software that draws physicians' attention to the regions in question, i.e. provide cues for sensemaking (Chapter 5).

Facilitate intuitive navigation: Navigation is one of the two key software interactions during the contouring task. Physicians should benefit from smooth navigation both within datasets (on different planes) as well as between datasets (Chapter 5).

Integrate intelligent Boolean operations with volumes: It is increasingly necessary to have the ability to include, exclude and combine certain regions (volumes), especially during the integration of medical knowledge into support software. However, Boolean operations without (semi-)automatic post-processing may result in illogical shapes. For instance, Vol_A *NOT* Vol_B is often expected to produce a single 3D volume, but if the contour boundaries are not precise there may be residual 'crumb' volumes. Undesirable outcomes of this kind could be eliminated by using more sophisticated algorithms (Chapters 4 and 5).

Incorporate automatic contouring: Computational algorithms are becoming more and more advanced and they need to be incorporated into the software. However, in this thesis we demonstrated that: (1) automation may bias physicians (Chapter 6); and (2) one size does not fit all, meaning that the software needs to be able to suggest algorithms depending on the contouring task and situation (Chapter 4).

Ensure usability of contouring tools: As a baseline, the usability of the contouring tools must be extremely good since usability problems can hinder physicians' thinking processes (Chapters 4 and 5).

Support comparison of contours: It is necessary to enable evaluation between contours, or contours on different medical images on different planes. Finally, 3D evaluation of contours could also be beneficial (Chapter 5).

REFERENCES

Abraham A, Petre M, Sharp H (2008) Information seeking: Sensemaking and interactions. In: CHI 2008 Sensemaking workshop. ISBN: 9781450300308

Al-Moteri MO, Symmons M, Plummer V, Cooper S (2017) Eye tracking to investigate cue processing in medical decision-making: A scoping review. Computers in Human Behavior 66:52–66 . doi: 10.1016/j.chb.2016.09.022

Anderson JR, Bothell D, Fincham JM, Anderson AR, Poole B, Qin Y (2011) Brain Regions Engaged by Part- and Whole-task Performance in a Video Game: A Model-based Test of the Decomposition Hypothesis. Journal of Cognitive Neuroscience 23:3983–3997 . doi: 10.1162/jocn_a_00033

Andre A, Wickens C (1995) When Users Want What's not Best for Them. Ergonomics in Design 8: . doi: 10.1177/106480469500300403

Annett J (2003) Hierarchical task analysis. In: Handbook of cognitive task design. pp 17–35

Aselmaa A, Goossens RHM, Freudenthal A (2013a) What is Sensemaking in the Context of External Radiotherapy Treatment Planning? In: DMD Europe 2013: Design of Medical Devices Conference - Europe Edition 2013. Delft University of Technology, Delft, The Netherlands

Aselmaa A, Goossens RHM, Laprie A, Ken S, Fechter T, Ramkumar A, Freudenthal A (2013b) Workflow Analysis Report. Delft University of Technology

Aselmaa A, Goossens RHM, Laprie A, Ramkumar A, Ken S, Freudenthal A (2013c) External radiotherapy treatment planning–situation today and perspectives for tomorrow. In: MASSOPTIER L, VIARD R (eds) Innovative imaging to improve radiotherapy treatments, 1st edn. Lulu Enterprises, pp 91–98, ISBN: 9781291604177

Aselmaa A, Goossens RHM, Rowland B, Laprie A, Song Y, Freudenthal A (2014) Medical Factors of Brain Tumor Delineation in Radiotherapy for Software Design. In: Ahram T, Karwowski W, Marek T (eds) 5th International Conference on Applied Human Factors and Ergonomics (AHFE). pp 4865–4875, ISBN: 978-1-4951-2093-0

Aselmaa A, Song Y, Goossens RHM (2016) User Interaction Visualization for Design Synthesis. In: Isenberg T, Sadlo F (eds) Poster Proceedings of the Eurographics Conference on Visualization (EuroVis 2016). The Eurographics Association, p 3 doi: http://dx.doi.org/10.2312/eurp.20161128

Aselmaa A, van Herk M, Laprie A, Nestle U, Götz I, Wiedenmann N, Schimek-Jasch T, Picaud F, Syrykh C, Cagetti L V, Jolnerovski M, Song Y, Goossens RHM (2017) Using a contextualized sensemaking model for interaction design: A case study of tumor contouring. Journal of Biomedical Informatics 65:145–158 . doi: 10.1016/j.jbi.2016.12.001

Avison DE, Lau F, Myers MD, Nielsen P a (1999) Action research. Communications of the

ACM 42:94–97 . doi: 10.1145/291469.291479

Barrett WA, Mortensen EN (1997) Interactive live-wire boundary extraction. Medical Image Analysis 1:331–341 . doi: 10.1016/S1361-8415(97)85005-0

Basu A, Blanning RW (2000) A Formal Approach to Workflow Analysis. Information Systems Research 11:17–36 . doi: 10.1287/isre.11.1.17.11787

Batumalai V, Holloway LC, Kumar S, Dundas K, Jameson MG, Vinod SK, Delaney GP (2016) Survey of image-guided radiotherapy use in Australia. Journal of Medical Imaging and Radiation Oncology. doi: 10.1111/1754-9485.12556

Bauer S, Wiest R, Nolte L-P, Reyes M (2013) A survey of MRI-based medical image analysis for brain tumor studies. Physics in Medicine and Biology 58:R97–R129 . doi: dx.doi.org/10.1088/0031-9155/58/13/R97

Behin A, Hoang-Xuan K, Carpentier AF, Delattre J-Y (2003) Primary brain tumours in adults. The Lancet 361:323–331 . doi: 10.1016/S0140-6736(03)12328-8

Bentzen SM (2005) Theragnostic imaging for radiation oncology: dose-painting by numbers. The Lancet Oncology 6:112–117 . doi: 10.1016/S1470-2045(05)01737-7

Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, Ten Haken RK, Yorke ED (2010) Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. International journal of radiation oncology, biology, physics 76:S3-9 . doi: 10.1016/j.ijrobp.2009.09.040

Bhavnani S, Clarkson G, Scholl M (2008) Collaborative search and sensemaking of patents. In: Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI ’08. ACM Press, New York, New York, USA, p 2799, ISBN: 978160558012X doi: 10.1145/1358628.1358764

Blanton M, Zhou J, Moro MM, Zhang D, Tsotras VJ, Halkidi M, Hainaut J-L, Tanin E, Korn FR, Pedersen TB, Böhm C, Plant C, Zhang Q, Strauss MJ, Jensen CS, Snodgrass R, Li N, Kantarcioglu M, Sebe N, Jaimes A, Dix A, Tompa FW (2009) Human-Computer Interaction. In: Encyclopedia of Database Systems. Springer US, Boston, MA, pp 1327–1331 doi: 10.1007/978-0-387-39940-9_192

Blazeby JM, Wilson L, Metcalfe C, Nicklin J, English R, Donovan JL (2006) Analysis of clinical decision-making in multi-disciplinary cancer teams. Annals of Oncology 17:457–460 . doi: 10.1093/annonc/mdj102

Blumenthal D (2009) Stimulating the Adoption of Health Information Technology. New England Journal of Medicine 360:1477–1479 . doi: 10.1056/NEJMp0901592

Bokstein F, Kovner F, Blumenthal DT, Ram Z, Templehoff H, Kanner AA, Corn BW (2008) A Common Sense Approach to Radiotherapy Planning of Glioblastoma Multiforme Situated in The Temporal Lobe. International Journal of Radiation Oncology*Biology*Physics 72:900–904 . doi: 10.1016/j.ijrobp.2008.01.053

Bossen C, Jensen LG (2014) How physicians “achieve overview.” In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing -

CSCW '14. ACM Press, New York, New York, USA, pp 257–268, ISBN: 9781450325400 doi: 10.1145/2531602.2531620

Brundage MD, Dixon PF, Mackillop WJ, Shelley WE, Hayter C, Paszat LF, Youssef YM, Robins JM, McNamee A, Cornell A (1999) A real-time audit of radiation therapy in a regional cancer center. International Journal of Radiation Oncology*Biology*Physics 43:115–124 . doi: 10.1016/S0360-3016(98)00368-X

Brunskill K, Nguyen TK, Boldt RG, Louie A V, Warner A, Marks LB, Palma DA (2017) Does Peer Review of Radiation Plans Affect Clinical Care? A Systematic Review of the Literature. International Journal of Radiation Oncology*Biology*Physics 97:27–34 . doi: 10.1016/j.ijrobp.2016.09.015

Burnet NG (2004) Defining the tumour and target volumes for radiotherapy. Cancer Imaging 4:153–161 . doi: 10.1102/1470-7330.2004.0054

Cancer Research UK uploader (2016) Diagram showing the parts of the brain CRUK 476.svg. https://commons.wikimedia.org/wiki/File%3ADiagram_showing_the_parts_of_the_brain_CRUK_476.svg. Accessed 30 Mar 2017

Castadot P, Lee JA, Geets X, Grégoire V (2010) Adaptive Radiotherapy of Head and Neck Cancer. Seminars in Radiation Oncology 20:84–93 . doi: 10.1016/j.semradonc.2009.11.002

Chan AJ, Islam MK, Rosewall T, Jaffray DA, Easty AC, Cafazzo JA (2012) Applying usability heuristics to radiotherapy systems. Radiotherapy and Oncology 102:142–147 . doi: 10.1016/j.radonc.2011.05.077

Chang EL, Akyurek S, Avalos T, Rebueno N, Spicer C, Garcia J, Famiglietti R, Allen PK, Chao KSC, Mahajan A, Woo SY, Maor MH (2007) Evaluation of Peritumoral Edema in the Delineation of Radiotherapy Clinical Target Volumes for Glioblastoma. International Journal of Radiation Oncology*Biology*Physics 68:144–150 . doi: 10.1016/j.ijrobp.2006.12.009

Chaouachi M, Jraidi I, Frasson C (2011) Modeling mental workload using EEG features for intelligent systems. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 50–61, ISBN: 9783642223617 doi: 10.1007/978-3-642-22362-4_5

Charette RN (2005) Why software fails [software failure. IEEE Spectrum 42:42–49 . doi: 10.1109/MSPEC.2005.1502528

Chen W (2007) Clinical applications of PET in brain tumors. Journal of Nuclear Medicine 48:1468–81 . doi: 10.2967/jnumed.106.037689

Chen Y., Alsakran J., Barlowe S., Yang J., Zhao Y. (2011) Supporting effective common ground construction in asynchronous collaborative visual analytics. In: VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings. Providence, RI, pp 101–110, ISBN: 9781467300131 doi: 10.1109/VAST.2011.6102447

Cheng J, Teevan J, Iqbal ST, Bernstein MS (2015) Break It Down. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15. ACM Press, New York, New York, USA, pp 4061–4064, ISBN: 9781450331456 doi: 10.1145/2702123.2702146

Colman H, Berkey BA, Maor MH, Groves MD, Schultz CJ, Vermeulen S, Nelson DF, Mehta MP, Yung WKA (2006) Phase II Radiation Therapy Oncology Group trial of conventional radiation therapy followed by treatment with recombinant interferon-β for supratentorial glioblastoma: Results of RTOG 9710. International Journal of Radiation Oncology*Biology*Physics 66:818–824 . doi: 10.1016/j.ijrobp.2006.05.021

Cooper A, Reimann R, Cronin D (2007) About Face 3: The essentials of interaction design. ISBN: 9780470084113

Corradini S, Niyazi M, Niemoeller OM, Li M, Roeder F, Eckel R, Schubert-Fritschle G, Scheithauer HR, Harbeck N, Engel J, Belka C (2015) Adjuvant radiotherapy after breast conserving surgery – A comparative effectiveness research study. Radiotherapy and Oncology 114:28–34 . doi: 10.1016/j.radonc.2014.08.027

Creak AL, Tree A, Saran F (2011) Radiotherapy Planning in High-grade Gliomas: a Survey of Current UK Practice. Clinical Oncology 23:189–198 . doi: 10.1016/j.clon.2010.11.003

Crum WR, Hartkens T, Hill DLG (2004) Non-rigid image registration: theory and practice. The British Journal of Radiology 77:S140–S153 . doi: 10.1259/bjr/25329214

Cuijpers CF, Moelker A, Varga E, Stappers PJ, Freudenthal A (2012) Improving image guidance in interventional radiology: Information lack in transjugular intrahepatic portosystemic shunt. In: EAM 2012: 30th European Annual Conference on Human Decision-Making and Manual Control

Dehnad H, Nederveen AJ, Heide UA va. der, van Moorselaar RJA, Hofman P, Lagendijk JJW (2003) Clinical feasibility study for the use of implanted gold seeds in the prostate as reliable positioning markers during megavoltage irradiation. Radiotherapy and Oncology 67:295–302 . doi: 10.1016/S0167-8140(03)00078-1

Delaney G, Jacob S, Featherstone C, Barton M (2005) The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines. Cancer 104:1129–1137 . doi: 10.1002/cncr.21324

Dervin B (1998) Sense-making theory and practice: an overview of user interests in knowledge seeking and use. Journal of Knowledge Management 2:36–46 . doi: 10.1108/13673279810249369

Dervin B (1983) An overview of sense-making research: Concepts, methods, and results to date. Int. Commun. Assoc. Annu. Meet. 1–14

Dervin B, Naumer CM (2009) Sense-Making. In: Encyclopedia of Library and Information Sciences, Third Edit. Taylor & Francis, pp 4696–4707, ISBN: 0-8493-9712-X

Diaz I, Boulanger P, Greiner R, Hoehn B, Rowe L, Murtha A (2013) An automatic brain tumor segmentation tool. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp 3339–3342, ISBN:

978-1-4577-0216-7 doi: 10.1109/EMBC.2013.6610256

DiDomenico A, Nussbaum MA (2005) Interactive effects of mental and postural demands on subjective assessment of mental workload and postural stability. In: Safety Science. pp 485–495, ISBN: 0925-7535 doi: 10.1016/j.ssci.2005.08.010

Dinkel J, Khalilzadeh O, Hintze C, Fabel M, Puderbach M, Eichinger M, Schlemmer HP, Thorn M, Heussel CP, Thomas M, Kauczor HU, Biederer J (2013) Inter-observer reproducibility of semi-automatic tumor diameter measurement and volumetric analysis in patients with lung cancer. Lung Cancer 82:76–82 . doi: 10.1016/j.lungcan.2013.07.006

Dolz J, Kirişli HA, Fechter T, Karnitzki S, Oehlke O, Nestle U, Vermandel M, Massoptier L (2016) Interactive contour delineation of organs at risk in radiotherapy: Clinical evaluation on NSCLC patients. Medical Physics 43:2569–2580 . doi: 10.1118/1.4947484

Donahue GM (2001) Usability and the bottom line. IEEE Software 18:31–37 . doi: 10.1109/52.903161

Dorn B, Guzdial M (2006) Graphic designers who program as informal computer science learners. Proceedings of the 2006 international workshop on Computing education research - ICER '06 127 . doi: 10.1145/1151588.1151608

Dou W, Jeong DH, Stukes F, Ribarsky W, Lipford HR, Chang R (2009) Recovering Reasoning Processes from User Interactions. IEEE Computer Graphics and Applications 29:52–61 . doi: 10.1109/MCG.2009.49

Dowsett RJ, Galvin JM, Cheng E, Smith R, Epperson R, Harris R, Henze G, Needham M, Payne R, Peterson MA, Skinner AL, Reynolds A (1992) Contouring structures for 3-dimensional treatment planning. International Journal of Radiation Oncology, Biology, Physics 22:1083–1088 . doi: 10.1016/0360-3016(92)90812-V

Drabik DM, MacKenzie MA, Fallone GB (2007) Quantifying Appropriate PTV Setup Margins: Analysis of Patient Setup Fidelity and Intrafraction Motion Using Post-Treatment Megavoltage Computed Tomography Scans. International Journal of Radiation Oncology*Biology*Physics 68:1222–1228 . doi: 10.1016/j.ijrobp.2007.04.007

Drevelegas A, Papanikolaou N (2011) Imaging Modalities in Brain Tumors. In: Imaging of Brain Tumors with Histological Correlations. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 13–33, ISBN: 9783540876489 doi: 10.1007/978-3-540-87650-2_2

Drew T, Evans K, Võ ML-H, Jacobson FL, Wolfe JM (2013) Informatics in Radiology: What Can You See in a Single Glance and How Might This Guide Visual Search in Medical Images? RadioGraphics 33:263–274 . doi: 10.1148/rg.331125023

Duffey RB, Saull JW (2002) Know the risk : learning from errors and accidents : safety and risk in today's technology. Butterworth-Heinemann, ISBN: 0080509738

Elstein A (1978) Medical Problem Solving: An Analysis of Clinical Reasoning. Harvard University Press, ISBN: 0674561252

Elstein AS, Schwartz A (2002) Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. British Medical Journal 324:729–732 . doi: 10.1136/bmj.324.7339.729

Endert A, Fiaux P, North C (2012) Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering. IEEE Transactions on Visualization and Computer Graphics 18:2879–2888 . doi: 10.1109/TVCG.2012.260

Eriksson H-E, Penker M (2000) Business modeling with UML : business patterns at work. John Wiley & Sons, ISBN: 9780471295518

Essig M, Hawighorst H, Schoenberg SO, Engenhart-Cabillic R, Fuss M, Debus J, Zuna I, Knopp M V, van Kaick G (1998) Fast fluid-attenuated inversion-recovery (FLAIR) MRI in the assessment of intraaxial brain tumors. Journal of Magnetic Resonance Imaging 8:789–798 . doi: 10.1002/jmri.1880080407

Fitton I, Cornelissen SAP, Duppen JC, Steenbakkers RJHM, Peeters STH, Hoebers FJP, Kaanders JHAM, Nowak PJCM, Rasch CRN, van Herk M (2011) Semi-automatic delineation using weighted CT-MRI registered images for radiotherapy of nasopharyngeal cancer. Medical Physics 38:4662–4666 . doi: 10.1118/1.3611045

Fotina I, Lutgendorf-Caucig C, Stock M, Potter R, Georg D (2012) Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 188:160–167 . doi: 10.1007/s00066-011-0027-6

Freudenthal A, Stüdeli T, Lamata P, Samset E (2011) Collaborative co-design of emerging multi-technologies for surgery. Journal of Biomedical Informatics 44:198–215 . doi: 10.1016/j.jbi.2010.11.006

Galy E, Cariou M, Mélan C (2012) What is the relationship between mental workload factors and cognitive load types? International Journal of Psychophysiology 83:269–275 . doi: 10.1016/j.ijpsycho.2011.09.023

Gebhardt BJ, Dobelbower MC, Ennis WH, Bag AK, Markert JM, Fiveash JB (2014) Patterns of failure for glioblastoma multiforme following limited-margin radiation and concurrent temozolomide. Radiation Oncology 9:130 . doi: 10.1186/1748-717X-9-130

Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G, Plevritis SK (2014) Glioblastoma Multiforme: Exploratory Radiogenomic Analysis by Using Quantitative Image Features. Radiology 273:131731 . doi: 10.1148/radiol.14131731

Ghose A, Lim G, Husain S (2010) Treatment for glioblastoma multiforme: current guidelines and Canadian practice. Curr Oncol 17:52–58

Grau C, Defourny N, Malicki J, Dunscombe P, Borras JM, Coffey M, Slotman B, Bogusz M, Gasparotto C, Lievens Y (2014) Radiotherapy equipment and departments in the European countries: Final results from the ESTRO-HERO survey. Radiother. Oncol. 112:155–164

Grégoire V, Jeraj R, Lee JA, O'Sullivan B (2012) Radiotherapy for head and neck tumours in

2012 and beyond: conformal, tailored, and adaptive? The Lancet Oncology 13:e292–e300 . doi: 10.1016/S1470-2045(12)70237-1

Gumienny R, Lindberg T, Meinel C (2011) Exploring the synthesis of information in design processes–opening the black-box. In: DS 68-6: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 6: Design Information and Knowledge, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011

Guo H, Gomez SR, Ziemkiewicz C, Laidlaw DH (2016) A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights. Visualization and Computer Graphics, IEEE Transactions on 22:51–60 . doi: 10.1109/TVCG.2015.2467613

Guzdial M, Santos PJ, Badre A, Hudson SE, Gray MH (1994) Analyzing and visualizing log files: A computational science of usability. http://hdl.handle.net/1853/3586

Hart SG (2006) Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Sage Publications, pp 904–908, ISBN: 1541-9312 doi: 10.1177/154193120605000909

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology 52:139–183 . doi: 10.1016/S0166-4115(08)62386-9

Haux R (2006) Health information systems – past, present, future. International Journal of Medical Informatics 75:268–281 . doi: 10.1016/j.ijmedinf.2005.08.002

Heckel F, Moltz JH, Tietjen C, Hahn HK (2013) Sketch-Based Editing Tools for Tumour Segmentation in 3D Medical Images. Computer Graphics Forum 32:144–157 . doi: 10.1111/cgf.12193

Hillstrom AP (2000) Repetition effects in visual search. Perception & psychophysics 62:800–817 . doi: 10.3758/BF03206924

Hollingsworth D, Hampshire U (1994) Workflow management coalition: The workflow reference model. Document Number TC00-1003

Huttenlocher DP, Klanderman GA, Rucklidge WJ (1993) Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 15:850–863 . doi: 10.1109/34.232073

International Commission on Radiation Units and Measurements (1999) ICRU Report 50. Prescribing, Recording, and Reporting Photon Beam Therapy. Bethesda, MD

International Ergonomics Association (IEA) (2017) What is Ergonomics? http://www.iea.cc/whats/index.html. Accessed 8 Feb 2017

International Organization for Standardization (2015) IEC 62366-1:2015, Medical devices - Part 1: Application of usability engineering to medical devices. http://www.iso.org/iso/catalogue_detail.htm?csnumber=63179. Accessed 8 Feb 2017

International Organization for Standardization (ISO) (2010) Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems (ISO 9241-210: 2010)

Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE (2010) A review of methods of analysis in contouring studies for radiation oncology. J Med Imaging Radiat Oncol 54:401–410 . doi: 10.1111/j.1754-9485.2010.02192.x

Johnson CM, Turley JP (2006) The significance of cognitive modeling in building healthcare interfaces. International Journal of Medical Informatics 75:163–172 . doi: 10.1016/j.ijmedinf.2005.06.003

Johnson CW (2006) Why did that happen? Exploring the proliferation of barely usable software in healthcare systems. Quality and Safety in Health Care 15:i76–i81 . doi: 10.1136/qshc.2005.016105

Jorritsma W, Cnossen F, Dierckx RA, Oudkerk M, van Ooijen PMA (2016) Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client. International journal of medical informatics 85:36–42 . doi: 10.1016/j.ijmedinf.2015.10.007

Joseph D (2004) The Practice of Design-Based Research: Uncovering the Interplay Between Design, Research, and the Real-World Context. Educational Psychologist 39:235–242 . doi: 10.1207/s15326985ep3904_5

Kaley T, Barani I, Chamberlain M, McDermott M, Panageas K, Raizer J, Rogers L, Schiff D, Vogelbaum M, Weber D, Wen P (2014) Historical benchmarks for medical therapy trials in surgery- and radiation-refractory meningioma: a RANO review. Neuro-Oncology 16:829–840 . doi: 10.1093/neuonc/not330

Kannampallil TG, Abraham J, Patel VL (2016) Methodological framework for evaluating clinical processes: A cognitive informatics perspective. Journal of Biomedical Informatics 64:342–351 . doi: 10.1016/j.jbi.2016.11.002

Kannampallil TG, Franklin A, Mishra R, Almoosa KF, Cohen T, Patel VL (2013) Understanding the nature of information seeking behavior in critical care: Implications for the design of health information technology. Artificial Intelligence in Medicine 57:21–29 . doi: 10.1016/j.artmed.2012.10.002

Karat C-M (1990) Cost-Benefit Analysis of Usability Engineering Techniques. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 34:839–843 . doi: 10.1177/154193129003401203

Karkavelas G, Tascos N (2011) Epidemiology, Histologic Classification, and Clinical Course of Brain Tumors. In: Imaging of Brain Tumors with Histological Correlations. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1–12 doi: 10.1007/978-3-540-87650-2_1

Karsh B-T, Weinger MB, Abbott PA, Wears RL (2010) Health information technology: fallacies and sober realities. Journal of the American Medical Informatics Association 17:617–623 . doi: 10.1136/jamia.2010.005637

Kasoju A, Petersen K, Mäntylä M V. (2013) Analyzing an automotive testing process with evidence-based software engineering. Information and Software Technology 55:1237–1259 . doi: 10.1016/j.infsof.2013.01.005

Kawamoto K, Houlihan C a, Balas EA, Lobach DF (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. British medical journal 5:1409–1416 . doi: 10.1136/bmj.38398.500764.8F

Ken S, Vieillevigne L, Franceries X, Simon L, Supper C, Lotterie J-A, Filleron T, Lubrano V, Berry I, Cassol E, Delannes M, Celsis P, Cohen-Jonathan EM, Laprie A (2013) Integration method of 3D MR spectroscopy into treatment planning system for glioblastoma IMRT dose painting with integrated simultaneous boost. Radiation Oncology 8:1 . doi: 10.1186/1748-717X-8-1

Kim S, Soergel D (2006) Selecting and measuring task characteristics as independent variables. Proceedings of the American Society for Information Science and Technology 42:n/a-n/a . doi: 10.1002/meet.14504201111

Kirrmann S, Gainey M, Röhner F, Hall M, Bruggmoser G, Schmucker M, Heinemann FE (2015) Visualization of data in radiotherapy using web services for optimization of workflow. Radiation Oncology 10: . doi: 10.1186/s13014-014-0322-3

Kirthi Koushik AS, Harish K, Avinash HU (2013) Principles of radiation oncology: a beams eye view for a surgeon. Indian journal of surgical oncology 4:255–62 . doi: 10.1007/s13193-013-0231-1

Kjeldskov J, Skov MB, Stage J (2010) A longitudinal study of usability in health care: Does time heal? International Journal of Medical Informatics 79: . doi: 10.1016/j.ijmedinf.2008.07.008

Klein G a., Phillips JK, Rall EL, Peluso D a. (2007) A data-frame theory of sensemaking. In: Hoffman RR (ed) Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making. Routledge, pp 113–155, ISBN: 9780203810088 doi: 10.4324/9780203810088

Klein G, Moon B, Hoffman R (2006a) Making Sense of Sensemaking 1: Alternative Perspectives. IEEE Intelligent Systems 21:70–73 . doi: 10.1109/MIS.2006.75

Klein, Moon, Hoffman (2006b) Making Sense of Sensemaking 2: A Macrocognitive Model. IEEE Intelligent Systems 21:88–92 . doi: 10.1109/MIS.2006.100

Kodagoda N, Pontis S, Simmie D, Attfield S, Wong BLW, Blandford A, Hankin C (2017) Using Machine Learning to Infer Reasoning Provenance From User Interaction Log Data. Journal of Cognitive Engineering and Decision Making 11:23–41 . doi: 10.1177/1555343416672782

Kolko J (2010) Abductive Thinking and Sensemaking: The Drivers of Design Synthesis. Design Issues 26:15–28 . doi: 10.1162/desi.2010.26.1.15

Kristiansen M, Obstfelder A, Lotherington AT (2015) Nurses' sensemaking of contradicting logics: An underexplored aspect of organisational work in nursing homes.

Scandinavian Journal of Management 31:330–337 . doi: 10.1016/j.scaman.2015.04.003

Kuijf HJ (2015) MeVisLab-Hausdorff-distance source code. https://github.com/hjkuijf/MeVisLab-Hausdorff-distance. Accessed 23 Mar 2017

Kushniruk AW, Patel VL (2004) Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of Biomedical Informatics 37:56–76 . doi: 10.1016/j.jbi.2004.01.003

Laprie A, Catalaa I, Cassol E, McKnight TR, Berchery D, Marre D, Bachaud J-M, Berry I, Moyal EC-J (2008) Proton Magnetic Resonance Spectroscopic Imaging in Newly Diagnosed Glioblastoma: Predictive Value for the Site of Postradiotherapy Relapse in a Prospective Longitudinal Study. International Journal of Radiation Oncology*Biology*Physics 70:773–781 . doi: 10.1016/j.ijrobp.2007.10.039

Lazar J, Jones A, Shneiderman B (2006) Workplace user frustration with computers: an exploratory investigation of the causes and severity. Behaviour & Information Technology 25:239–251 . doi: 10.1080/01449290500196963

Ledley RS, Lusted LB (1959) Reasoning Foundations of Medical Diagnosis: Symbolic logic, probability, and value theory aid our understanding of how physicians reason. Science 130:9–21 . doi: 10.1126/science.130.3366.9

Leibel SA, Fuks Z, Zelefsky MJ, Wolden SL, Rosenzweig KE, Alektiar KM, Hunt MA, Yorke ED, Hong LX, Amols HI (2002) Intensity-Modulated Radiotherapy. The Cancer journal 8:164–176

Leykum LK, Chesser H, Lanham HJ, Carla P, Palmer R, Ratcliffe T, Reisinger H, Agar M, Pugh J (2015) The Association Between Sensemaking During Physician Team Rounds and Hospitalized Patients' Outcomes. Journal of General Internal Medicine 30:1821–1827 . doi: 10.1007/s11606-015-3377-4

Lim JY, Leech M (2016) Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. Acta Oncologica 55:799–806 . doi: 10.3109/0284186X.2016.1173723

Lindsay PH, Norman DA (1972) Problem solving. In: Lindsay PH, Norman DA (eds) Human Information Processing. Elsevier, pp 500–525, ISBN: 978-0-12-450950-4 doi: 10.1016/B978-0-12-450950-4.50019-6

Ling CC, Humm J, Larson S, Amols H, Fuks Z, Leibel S, Koutcher JA (2000) Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. International Journal of Radiation Oncology*Biology*Physics 47:551–560 . doi: 10.1016/S0360-3016(00)00467-3

Louie A V, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, Ahmad B, Aivas I, Gaede S (2010) Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era. Radiother Oncol 95:166–171 . doi: 10.1016/j.radonc.2009.12.028

Macdonald DR, Cascino TL, Schold SC, Cairncross JG (1990) Response criteria for phase II studies of supratentorial malignant glioma. Journal of Clinical Oncology 8:1277–1280 . doi: 10.1200/JCO.1990.8.7.1277

Mackenzie J, Graham G, Olivotto IA (2016) Peer Review of Radiotherapy Planning: Quantifying Outcomes and a Proposal for Prospective Data Collection. Clinical Oncology 28:e192–e198 . doi: 10.1016/j.clon.2016.08.012

Mackie TR, Holmes T, Swerdloff S, Reckwerdt P, Deasy JO, Yang J, Paliwal B, Kinsella T (1993) Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy. Medical Physics 20:1709–1719 . doi: 10.1118/1.596958

Mader EC (2015) What is the difference between T1 and T2 imaging in MRI? - Quora. https://www.quora.com/What-is-the-difference-between-T1-and-T2-imaging-in-MRI. Accessed 29 Mar 2017

Malhotra Y (2001) Expert systems for knowledge management: crossing the chasm between information processing and sense making. Expert Systems with Applications 20:7–16 . doi: 10.1016/S0957-4174(00)00045-2

Malý I, Slavík P (2007) Towards Visual Analysis of Usability Test Logs Using Task Models. In: Coninx K, Luyten K, Schneider KA (eds) Task Models and Diagrams for Users Interface Design. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 24–38, ISBN: 978-3-540-70816-2 doi: 10.1007/978-3-540-70816-2_3

Mamykina L, Nakikj D, Elhadad N (2015a) Collective Sensemaking in Online Health Forums. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15. ACM Press, New York, New York, USA, pp 3217–3226, ISBN: 9781450331456 doi: 10.1145/2702123.2702566

Mamykina L, Smaldone AM, Bakken SR (2015b) Adopting the sensemaking perspective for chronic disease self-management. Journal of Biomedical Informatics 56:406–417 . doi: 10.1016/j.jbi.2015.06.006

Mannion M, Keepence B (1995) SMART requirements. ACM SIGSOFT Software Engineering Notes 20:42–47 . doi: 10.1145/224155.224157

Manzey D, Reichenbach J, Onnasch L (2012) Human Performance Consequences of Automated Decision Aids. Journal of Cognitive Engineering and Decision Making 6:57–87 . doi: 10.1177/1555343411433844

Marks LB, Adams RD, Pawlicki T, Blumberg AL, Hoopes D, Brundage MD, Fraass BA (2013) Enhancing the role of case-oriented peer review to improve quality and safety in radiation oncology: Executive summary. Practical Radiation Oncology 3:149–156 . doi: 10.1016/j.prro.2012.11.010

Mason WP, Maestro RD, Eisenstat D, Forsyth P, Fulton D, Laperriere N, Macdonald D, Perry J, Thiessen B (2007) Canadian recommendations for the treatment of glioblastoma multiforme. Curr Oncol 14:110–117 . doi: 10.3747/co.v18i3.755

Mattson MP (2014) Superior pattern processing is the essence of the evolved human brain. Frontiers in Neuroscience 8: . doi: 10.3389/fnins.2014.00265

MeVis Medical Solutions AG (2016) MeVisLab – development environment for medical image processing and visualization.

Mirel B (2004) Interaction design for complex problem solving: Developing useful and usable software. Morgan Kaufmann, ISBN: 1558608311

Moeller BJ, Dewhirst MW (2006) HIF-1 and tumour radiosensitivity. British Journal of Cancer 95:1–5 . doi: 10.1038/sj.bjc.6603201

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 24:794–803 . doi: 10.1007/s10278-010-9341-2

Munoz MI, Mollo V, Barcellini F, Nascimento A (2011) Design of a safe computer-supported cooperative environment in radiotherapy. In: Proceedings of the 29th Annual European Conference on Cognitive Ergonomics - ECCE ’11. ACM Press, New York, New York, USA, p 197, ISBN: 9781450310291 doi: 10.1145/2074712.2074751

National Health Service (UK) (2015) Radiotherapy. http://www.nhs.uk/Conditions/radiotherapy/Pages/introduction.aspx. Accessed 16 Feb 2017

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall Englewood Cliffs, NJ, ISBN: 978-0134454030

Nielsen J (2012) Usability 101: Introduction to Usability. https://www.nngroup.com/articles/usability-101-introduction-to-usability/. Accessed 2 Feb 2017

Nielsen J (1995) 10 Usability Heuristics for User Interface Design. http://www.nngroup.com/articles/ten-usability-heuristics/

Nijkamp J, Pos FJ, Nuver TT, de Jong R, Remeijer P, Sonke J-J, Lebesque J V. (2008) Adaptive Radiotherapy for Prostate Cancer Using Kilovoltage Cone-Beam Computed Tomography: First Clinical Results. International Journal of Radiation Oncology*Biology*Physics 70:75–82 . doi: 10.1016/j.ijrobp.2007.05.046

Niyazi M, Brada M, Chalmers AJ, Combs SE, Erridge SC, Fiorentino A, Grosu AL, Lagerwaard FJ, Minniti G, Mirimanoff R-O, Ricardi U, Short SC, Weber DC, Belka C (2016) ESTRO-ACROP guideline “target delineation of glioblastomas.” Radiotherapy and Oncology. doi: 10.1016/j.radonc.2015.12.003

Njeh CF (2008) Tumor delineation: The weakest link in the search for accuracy in radiotherapy. Journal of medical physics / Association of Medical Physicists of India 33:136–40 . doi: 10.4103/0971-6203.44472

Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP (2002) Time Course of Perception and Decision Making During Mammographic Interpretation. American Journal of Roentgenology 179:917–923 . doi: 10.2214/ajr.179.4.1790917

Norman D (1986) Cognitive engineering. In: User Centered System Design: New

Perspectives on Human-Computer Interaction. pp 31–61, ISBN: 0898597811

Noyes JM, Bruneau DPJ (2007) A self-analysis of the NASA-TLX workload measure. Ergonomics 50:514–519 . doi: 10.1080/00140130701235232

Nutting C, Dearnaley DP, Webb S (2000) Intensity modulated radiation therapy: a clinical review. The British Journal of Radiology 73:459–469 . doi: 10.1259/bjr.73.869.10884741

Olabarriaga S., Smeulders AW. (2001) Interaction in the segmentation of medical images: A survey. Medical Image Analysis 5:127–142 . doi: 10.1016/S1361-8415(00)00041-4

Öllinger M, Goel V (2010) Problem Solving. In: Towards a Theory of Thinking. Springer, pp 3–21, ISBN: 3642031285 doi: 10.1007/978-3-642-03129-8_1

Olsen LA, Robinson CG, He GR, Wooten HO, Yaddanapudi S, Mutic S, Yang D, Moore KL (2014) Automated radiation therapy treatment plan workflow using a commercial application programming interface. Practical Radiation Oncology 4:358–367 . doi: 10.1016/j.prro.2013.11.007

Osheroff JA, Pifer EA, Sittig DF, Jenders RA, Teich JM (2004) Clinical Decision Support Implementers' Workbook. ISBN: 0-9725371-8-X 230

Pantanowitz L, Parwani A, Tseytlin E, Mello-Thoms C (2012) Analysis of slide exploration strategy of cytologists when reading digital slides. In: Abbey CK, Mello-Thoms CR (eds) SPIE Proceedings, Medical Imaging. International Society for Optics and Photonics, p 83180O doi: 10.1117/12.912498

Pantazi S V, Arocha JF, Moehr JR (2004) Case-based medical informatics. BMC medical informatics and decision making 4:19 . doi: 10.1186/1472-6947-4-19

Pantazi S V, Kushniruk A, Moehr JR (2006) The usability axiom of medical information systems. International Journal of Medical Informatics 75:829–839 . doi: 10.1016/j.ijmedinf.2006.05.039

Parascandolo P, Cesario L, Vosilla L, Pitikakis M, Viano G (2013) Smart Brush: A real time segmentation tool for 3D medical images. In: 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, pp 689–694, ISBN: 978-953-184-194-8 doi: 10.1109/ISPA.2013.6703826

Parasuraman R, Manzey DH (2010) Complacency and bias in human use of automation: an attentional integration. Human factors 52:381–410 . doi: 10.1177/0018720810376055

Parker W, Patrocinio H (2005) Chapter 7 Clinical treatment planning in external photon beam radiotherapy. In: Radiation oncology physics: A handbook for teachers and students. pp 219–272, ISBN: 9201073046

Patel VL, Kannampallil TG (2015) Cognitive informatics in biomedicine and healthcare. Journal of Biomedical Informatics 53:3–14 . doi: 10.1016/j.jbi.2014.12.007

Patel VL, Kaufman DR (2006) Cognitive science and biomedical informatics. In: Biomedical informatics. Springer, pp 133–185, ISBN: 0387289860

Patterson RE, Blaha LM, Grinstein GG, Liggett KK, Kaveney DE, Sheldon KC, Havig PR, Moore JA (2014) A human cognition framework for information visualization. Computers & Graphics 42:42–58 . doi: 10.1016/j.cag.2014.03.002

Paul SA., Reddy MC., DeFlitch CJ. (2008) Information and communication tools as aids to collaborative sensemaking. In: Conference on Human Factors in Computing Systems - Proceedings. Florence, pp 3105–3110 doi: 10.1145/1358628.1358815

Paul SA, Reddy MC (2010) Understanding together. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW '10. ACM Press, New York, New York, USA, p 321, ISBN: 9781605587950 doi: 10.1145/1718918.1718976

Peleg M (2013) Computer-interpretable clinical guidelines: A methodological review. Journal of Biomedical Informatics 46:744–763 . doi: 10.1016/j.jbi.2013.06.009

Pereira GC, Traughber M, Muzic RF (2014) The role of imaging in radiation therapy planning: past, present, and future. BioMed research international 2014:231090 . doi: 10.1155/2014/231090

Pham JC, Aswani MS, Rosen M, Lee H, Huddle M, Weeks K, Pronovost PJ (2012) Reducing Medical Errors and Adverse Events. Annual Review of Medicine 63:447–463 . doi: 10.1146/annurev-med-061410-121352

Pirolli P, Card S (1999) Information foraging. Psychological Review 106:643–675 . doi: 10.1037/0033-295X.106.4.643

Pirolli P, Card S (2005) The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In: Proceedings of international conference on intelligence analysis

Pirolli P, Russell D (2011) Introduction to this Special Issue on Sensemaking. Human-Computer Interaction 26:1–8 . doi: 10.1080/07370024.2011.556557

Poon EG, Kachalia A, Puopolo AL, Gandhi TK, Studdert DM (2012) Cognitive Errors and Logistical Breakdowns Contributing to Missed and Delayed Diagnoses of Breast and Colorectal Cancers: A Process Analysis of Closed Malpractice Claims. Journal of General Internal Medicine 27:1416–1423 . doi: 10.1007/s11606-012-2107-4

Prabhakar R, Haresh K, Laviraj M, Julka P, Rath G, Manoharan N, Rajendran M (2011) A study on the tumor volume computation between different 3D treatment planning systems in radiotherapy. Journal of Cancer Research and Therapeutics 7:168 . doi: 10.4103/0973-1482.82917

Proctor RW, Wu KPL (2007) Human Information Processing: An Overview for Human-Computer Interaction. Sears, A., Jacko, J. A. In: The Human-Computer Interaction Handbook. pp 43–62, ISBN: 9781410615862

Purdy JA (2004) Current ICRU definitions of volumes: limitations and future directions. Seminars in Radiation Oncology 14:27–40 . doi: 10.1053/j.semradonc.2003.12.002

Ramkumar A, Dolz J, Kirisli HA, Adebahr S, Schimek-Jasch T, Nestle U, Massoptier L, Varga E, Stappers PJ, Niessen WJ, Song Y (2016) User Interaction in Semi-Automatic Segmentation of Organs at Risk: a Case Study in Radiotherapy. Journal of Digital Imaging 29:264–277 . doi: 10.1007/s10278-015-9839-8

Ramkumar A, Song Y, Varga E, Niessen WJ, Laprie A, Rowland B, Freudenthal A (2014) Comparison of Heuristic Evaluation and Think Aloud Methods: A Study in Radiotherapy Contouring Software. In: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care. SAGE PublicationsSage India: New Delhi, India, pp 230–237 doi: 10.1177/2327857914031038

Ramkumar A, Varga E, Laprie A, Niessen WJ, Freudenthal A (2013a) A pilot study of Pen-Tablet interaction in radiotherapy contouring using orthogonal and non-orthogonal views. In: Design of Medical Devices Conference--Europe (DMD'13--Delft). pp 44–50

Ramkumar A, Varga E, Niessen WJ, Freundethal A (2013b) Exploring input devices for contouring in external radiotherapy. In: Innovative imaging to improve radiotherapy treatments, Lulu Enterprises Inc Ed. pp 29–34, ISBN: 978-1-291-60417-7

Reddy MC, Jansen BJ (2008) A model for understanding collaborative information behavior in context: A study of two healthcare teams. Information Processing & Management 44:256–273 . doi: 10.1016/j.ipm.2006.12.010

Roberge D, Skamene T, Turcotte RE, Powell T, Saran N, Freeman C (2011) Inter- and intra-observer variation in soft-tissue sarcoma target definition. Cancer/Radiothérapie 15:421–425 . doi: 10.1016/j.canrad.2011.03.006

Rosman B, Ramamoorthy S, Mahmud MMH, Kohli P (2014) On user behaviour adaptation under interface change. International Conference on Intelligent User Interfaces, Proceedings IUI 273–278 . doi: 10.1145/2557500.2557535

Russell DM, Furnas G, Stefik M, Card SK, Pirolli P (2008) Sensemaking. In: Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08. ACM Press, New York, New York, USA, p 3981, ISBN: 978160558012X doi: 10.1145/1358628.1358972

Russell DM, Pirolli P, Furnas G, Card SK, Stefik M (2009) Sensemaking workshop CHI 2009. In: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09. ACM Press, New York, New York, USA, p 4751, ISBN: 9781605582474 doi: 10.1145/1520340.1520732

Russell DM, Stefik MJ, Pirolli P, Card SK (1993) The cost structure of sensemaking. In: Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93. ACM Press, New York, New York, USA, pp 269–276, ISBN: 0897915755 doi: 10.1145/169059.169209

Sanders EB-N, Stappers PJ (2014) Probes, toolkits and prototypes: three approaches to making in codesigning. Codesign-International Journal of Cocreation in Design and the Arts 10:5–14 . doi: 10.1080/15710882.2014.888183

Saposnik G, Redelmeier D, Ruff CC, Tobler PN (2016) Cognitive biases associated with medical decisions: a systematic review. BMC medical informatics and decision making 16:138 . doi: 10.1186/s12911-016-0377-1

Sass L, Oxman R (2006) Materializing design: The implications of rapid prototyping in digital design. Design Studies 27:325–355 . doi: 10.1016/j.destud.2005.11.009

Sauer J, Seibel K, Rüttinger B (2010) The influence of user expertise and prototype fidelity in usability tests. Applied Ergonomics 41:130–140 . doi: 10.1016/j.apergo.2009.06.003

Schaffner B, Pedroni E (1998) The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power. Physics in Medicine and Biology 43:1579–1592 . doi: 10.1088/0031-9155/43/6/016

Sedlmair M, Meyer M, Munzner T (2012) Design study methodology: Reflections from the trenches and the stacks. Visualization and Computer Graphics, IEEE Transactions on 18:2431–2440

Sharma N (2009) Sensemaking handoff: When and how? Proceedings of the American Society for Information Science and Technology 45:1–12 . doi: 10.1002/meet.2008.1450450234

Shi Y, Park T, Ruiz N, Taib R, Choi EHC, Chen F (2007) Galvanic Skin Response (GSR) as an Index of Cognitive Load. CHI EA ’07 CHI ’07 Extended Abstracts on Human Factors in Computing Systems 2651–2656 . doi: 10.1145/1240866.1241057

Sisterson J (2005) Ion beam therapy in 2004. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 241:713–716 . doi: 10.1016/j.nimb.2005.07.121

Skitka LJ, Mosier KL, Burdick M, State SJ, Ames N (1999) Does automation bias decision-making? International Journal of Human-Computer Studies 51:991–1006 . doi: 10.1006/ijhc.1999.0252

Slashme (2014) File:PlanesOfSection.jpg - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:PlanesOfSection.jpg. Accessed 29 Mar 2017

Slotman BJ, Leer JWH (2003) Infrastructure of radiotherapy in the Netherlands: Evaluation of prognoses and introduction of a new model for determining the needs. Radiotherapy and Oncology 66:345–349 . doi: 10.1016/S0167-8140(03)00025-2

Smith LK, Pope C, Botha JL (2005) Patients’ help-seeking experiences and delay in cancer presentation: a qualitative synthesis. The Lancet 366:825–831 . doi: 10.1016/S0140-6736(05)67030-4

Song Y, Hoeksema J, Ramkumar A, Molenbroek JFM (2017) A landmark based 3D parametric foot model for footwear customization. International Journal of the Digital Human (accepted)

Stanton N, Hedge A, Brookhuis K, Salas E, Hendrick H (2005) Handbook of Human Factors and Ergonomics Methods. ISBN: 0203489926

Stead WW, Lin HS, National Research Council (US) Committee on Engaging the Informatics, Computer Science Research Community in Health Care (2009) Computational Technology for Effective Health Care. National Academies Press, Washington, D.C., ISBN: 978-0-309-13050-9

Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KEI, Zijp L, Uitterhoeve ALJ, Rodrigus PTR, Kramer GWP, Bussink J, Jaeger K De, Belderbos JSA, Hart AAM, Nowak PJCM, Herk M van, Rasch CRN (2005) Observer variation in target volume delineation of lung cancer related to radiation oncologist–computer interaction: A “Big Brother” evaluation. Radiotherapy and Oncology 77:182–190 . doi: 10.1016/j.radonc.2005.09.017

Stolterman E, Wiberg M (2010) Concept-Driven Interaction Design Research. Human-Computer Interaction 25:95–118 . doi: 10.1080/07370020903586696

Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff R-O (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. The Lancet Oncology 10:459–466 . doi: 10.1016/S1470-2045(09)70025-7

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO (2005) Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England Journal of Medicine 352:987–996 . doi: 10.1056/NEJMoa043330

Sykes J (2014) Reflections on the current status of commercial automated segmentation systems in clinical practice. Journal of Medical Radiation Sciences 61:131–134 . doi: 10.1002/jmrs.65

Tan H, Chen L, Guan Y, Lin X (2011) Comparison of MRI, F-18 FDG, and 11C-Choline PET/CT for Their Potentials in Differentiating Brain Tumor Recurrence From Brain Tumor Necrosis Following Radiotherapy. Clinical Nuclear Medicine 36:978–981 . doi: 10.1097/RLU.0b013e31822f68a6

Teixeira L, Ferreira C, Santos BS (2012) User-centered requirements engineering in health information systems: A study in the hemophilia field. Computer Methods and Programs in Biomedicine 106:160–174 . doi: 10.1016/j.cmpb.2010.10.007

The Royal College of Radiologists, Society and College of Radiographers, Institute of Physics and Engineering in Medicine, National Patient Safety Agency, British Institute of Radiology (2008) Towards Safer Radiotherapy. ISBN: 978-1-905034-25-3

van den Haak M, De Jong M, Jan Schellens P (2003) Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. Behaviour & Information Technology 22:339–351 . doi: 10.1080/0044929031000

van Herk M (2004) Errors and margins in radiotherapy. Semin Radiat Oncol 14:52–64 . doi: 10.1053/j.semradonc.2003.10.003

Varga E, Pattynama PMT, Freudenthal A (2013) Manipulation of mental models of anatomy in interventional radiology and its consequences for design of human–computer interaction. Cognition, Technology & Work 15:457–473 . doi: 10.1007/s10111-012-0227-6

Vicente KJ (2010) The Human Factor: Revolutionizng the Way People Live with Technology. ISBN: 0307366146

Vicente KJ (2002) Ecological Interface Design: Progress and Challenges. Human Factors: The Journal of the Human Factors and Ergonomics Society 44:62–78 . doi: 10.1518/0018720024494829

Vieira B, Hans EW, van Vliet-Vroegindeweij C, van de Kamer J, van Harten W (2016) Operations research for resource planning and -use in radiotherapy: a literature review. BMC Medical Informatics and Decision Making 16:149 . doi: 10.1186/s12911-016-0390-4

Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I (2011) National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability. International journal of medical informatics 80:708–725 . doi: http://dx.doi.org/10.1016/j.ijmedinf.2011.06.010

Vorwerk H, Beckmann G, Bremer M, Degen M, Dietl B, Fietkau R, Gsänger T, Hermann RM, Alfred Herrmann MK, Höller U, van Kampen M, Körber W, Maier B, Martin T, Metz M, Richter R, Siekmeyer B, Steder M, Wagner D, Hess CF, Weiss E, Christiansen H (2009) The delineation of target volumes for radiotherapy of lung cancer patients. Radiotherapy and Oncology 91:455–460 . doi: 10.1016/j.radonc.2009.03.014

Vorwerk H, Zink K, Schiller R, Budach V, Böhmer D, Kampfer S, Popp W, Sack H, Engenhart-Cabillic R (2014) Protection of quality and innovation in radiation oncology: The prospective multicenter trial the German Society of Radiation Oncology (DEGRO-QUIRO study). Strahlentherapie und Onkologie 190:433–443 . doi: 10.1007/s00066-014-0634-0

Wee CW, Sung W, Kang H-C, Cho KH, Han TJ, Jeong B-K, Jeong J-U, Kim H, Kim IA, Kim JH, Kim SH, Kim S, Lee DS, Lee MY, Lim DH, Park HL, Suh C-O, Yoon SM, Kim IH (2016) Evaluation of variability in target volume delineation for newly diagnosed glioblastoma: a multi-institutional study from the Korean Radiation Oncology Group. Radiation Oncology 10:137 . doi: 10.1186/s13014-015-0439-z

Weersink RA (2016) Chapter 3 – Image Fusion and Visualization. In: Bioengineering for Surgery. pp 29–58, ISBN: 9780081001233 doi: 10.1016/B978-0-08-100123-3.00003-8

Weick KE (1995a) Sensemaking on organizations. ISBN: 080397177X

Weick KE (1995b) Sensemaking in organizations. SAGE Publications, Inc, ISBN: 080397177X

Weiss E, Hess CF (2003) The Impact of Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) Definition on the Total Accuracy in Radiotherapy. Strahlentherapie und Onkologie 179:21–30 . doi: 10.1007/s00066-003-0976-5

Weltens C, Menten J, Feron M, Bellon E, Demaerel P, Maes F, Van den Bogaert W, van der Schueren E (2001) Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. Radiotherapy and Oncology 60:49–59 . doi: 10.1016/S0167-8140(01)00371-1

Wesley D, Dau LA (2017) Complacency and Automation Bias in the Enbridge Pipeline Disaster. Ergonomics in Design 25:17–22 . doi: 10.1177/1064804616652269

Whitfield GA, Price P, Price GJ, Moore CJ (2013) Automated delineation of radiotherapy volumes: are we going in the right direction? Br J Radiol 86:20110718 . doi: doi:10.1259/bjr.20110718

Wilms G, Lammens M, Marchal G, Demaerel P, Verplancke J, Van Calenbergh F, Goffin J, Plets C, Baert AL (1991) Prominent dural enhancement adjacent to nonmeningiomatous malignant lesions on contrast-enhanced MR images. American Journal of Neuroradiology 12:761–764

Winkel D, Bol GH, van Asselen B, Hes J, Scholten V, Kerkmeijer LGW, Raaymakers BW (2016) Development and clinical introduction of automated radiotherapy treatment planning for prostate cancer. Physics in Medicine and Biology 61:8587–8595 . doi: 10.1088/1361-6560/61/24/8587

Xing L, Thorndyke B, Schreibmann E, Yang Y, Li T-F, Kim G-Y, Luxton G, Koong A (2006) Overview of image-guided radiation therapy. Medical Dosimetry 31:91–112 . doi: 10.1016/j.meddos.2005.12.004

Yancik R, Ries LAG (2004) Cancer in Older Persons: An International Issue in an Aging World. Seminars in Oncology 31:128–136 . doi: 10.1053/j.seminoncol.2003.12.024

Yusof MM, Papazafeiropoulou A, Paul RJ, Stergioulas LK (2008) Investigating evaluation frameworks for health information systems. International Journal of Medical Informatics 77:377–385 . doi: 10.1016/j.ijmedinf.2007.08.004

Zabramski S (2011) Careless touch. In: Proceedings of the 23rd Australian Computer-Human Interaction Conference on - OzCHI ’11. ACM Press, New York, New York, USA, pp 329–332, ISBN: 9781450310901 doi: 10.1145/2071536.2071588

Zhang P, Soergel D (2014) Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. Journal of the Association for Information Science and Technology 65:1733–1756 . doi: 10.1002/asi.23125

Zhao F, Xie X (2013) An Overview of Interactive Medical Image Segmentation. Annals of the BMVA 2013:1–22

Zhu Y (2013) Towards More Desirable Segmentation via User Interactions. Lehigh University

(2017a) Objectives (SUMMER). http://summer-project.eu/about-summer/objectives/. Accessed 5 Mar 2017

(2010) Radiation Therapy for Cancer - National Cancer Institute. https://www.cancer.gov/about-cancer/treatment/types/radiation-therapy/radiation-fact-sheet. Accessed 4 Feb 2017

(2015) Electrons. http://www.myradiotherapy.com/general/treatment/Treatment_Machines/electrons/electrons_radiotherapy.html. Accessed 4 Feb 2017

(2013a) Phase III Study Comparing 2 Brain Conformational Radiotherapy in Combination With Chemotherapy in the Treatment of Glioblastoma (SPECTRO GLIO). https://clinicaltrials.gov/ct2/show/NCT01507506. Accessed 30 Mar 2017

(2008) RADIOTHERAPY RISK PROFILE. World Health Organization

(2013b) VASARI Features v3.0. https://docs.google.com/spreadsheet/viewform?formkey=dHIxSC1vOXZQUGZfV19KUkhNbmVYRHc6MA. Accessed 6 Feb 2017

(2013c) VASARI Research Project. https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project. Accessed 6 Feb 2017

(2012) About SUMMER. http://summer-project.eu/about-summer/. Accessed 30 Mar 2017

(2017b) Scopus. www.scopus.com. Accessed 30 Mar 2017

ACKNOWLEDGEMENTS

It has been a long journey with its ups and downs and along the way there have been many people who have made it possible. I am not one with many words, but here are a few.

First of all, I would like to thank my supervisors: Richard, Wolf, Adinda. Thanks to you, I have become an independent researcher. Our conversations, though challenging at times, pushed me to explore further and thus become better. Richard, your positive attitude and optimism helped me in so many ways. Wolf, you become part of the team later and I can only imagine how difficult it must have been. But your dedication, hard work, and kind words are invaluable for me. Dear Adinda, you gave me this opportunity and helped me get started with research. I am deeply sad that you are no longer with us, but also grateful that I had the chance to work with you. Another word of thanks belongs to Sylvia, my PhD mentor – even though we had little contact, knowing that I could always turn to you if there was a need, meant a lot.

Doctoral committee members, Prof.dr. MD A. Laprie, Prof.dr. MD U. Nestle, Prof.dr.ir. J.H. Eggen , Prof.dr.ir. J.M.P. Geraedts, Prof.dr. P.J. Stappers, thank you for reading the dissertation and providing valuable feedback on the draft.

This research was part of the SUMMER project, and all the member of it have my biggest thanks. Your advices and comments have been valuable. I would like to thank the consortium project leaders: Dr. D. Gibon, Mr. Laurent, Prof.dr. MD A. Laprie, Prof.dr. MD U. Nestle, Dr. MD U. Sabatini, Dipl.-Math.in Dr.in Katja Bühler, and Prof. Dr. W. Birkfellner, our conversations were always inspirational. I would also like to thank the staff of Institut Claudius Regaud, Oncopole, Toulouse, France and University Medical Centre, Freiburg, Germany for welcoming me to conduct research there and giving their time to participate in my studies.

I would also like to thank all the other researchers of the SUMMER project – Anjana, Jose, Hortense, Andrea, Ben, Solea, Tobias, Montse, Noora, Andac, Miguel, Matthias, Miro, and Frida. We had joy, we had fun… I'm waiting for that reunion.

My special thanks goes to Prof. Dr. M. van Herk. I found our meetings always motivational. And thanks to you, this research in this form was possible. I would also like to express my appreciations to Mr. J. C. Duppen, who

developed the original “Big Brother” software, and Mr. L. Ploeger, for technical support during software prototype development. Furthermore, I would also like to thank the staff of Netherlands Cancer Institute for welcoming me there as a guest.

There were a few people who helped my life in a non-research ways. Thank you Csilla, Mariska, Selina, and Sara for the help with all the financial-bureaucratic-administrative questions. And thank you Erica for helping me figure out my maternity leave.

I would like to thank also all of my colleagues. I like to thank Edit and Cecile. Starting my PhD with you as my colleagues made a huge difference. Prang, it is a pity that we had so little roommate time together, but it was valued. Bahar, we shared happiness and the frustrations together, and I’m looking forwards for quieter times to enjoy Delft together. Jotte, I’m so happy we met early into the PhD – our chats have been great. Holly, I wonder where you will be next, but in any case, don’t be a stranger. And thanks to everyone else I ran into in the coffee corner, or in the elevator, or elsewhere – those chats gave me a sense of belonging.

I would also like to thank all my friends near and far. Marju, thank you for being there and listening and sharing from distance. And Kairi, you were always there for me. Laura, living in Netherlands brought us together and I’m glad it did. Thank you for listening me vent at times and also pushing me to do more sports ☺. Darina and your funny discoveries – thank you for stretching my world. Wieke, thank you for being around and being a listening ear. Lucien & Cristina, thank you for the shelter when I moved here. It was great to have you as almost neighbors for a year. Mik, I have appreciated your company and cheerfulness. Gert, thank you for teaching me how to remove anchor screws. vTeele, thanks for those fun & fond memories, yatzy and all. Paal, your winter cabin will see us again. Marius & Inez, thank you for joining our events more than once despite the distance. Anca, seeing you has always been a pleasure and a joy. And thank you everyone else, for visiting us, or simply being a virtual friend.

I would like to thank my new colleagues for the patience and understanding at this time of finishing the thesis and getting ready for the defense.

In the end, I would like to thank all the members of my family. I would like to thank my family by marriage, Mária, László, Gábor, and Gréti for all the support and help - *nagyon szépen köszönöm*! And then few words in Estonian.

Aitäh Marko ja Terje et te süstisite minuses usku, et ma saan kõigega hakkama – suur suur aitäh! Ivor, Thomas, Kristi, Sebastian ja Melissa, Merly ja Ronan, ja ka kõik teised. Suur aitäh teile kõigile, et te olite mulle toeks lähedalt või kaugelt.

A big thank you goes also to my little girl Karola. You being born taught me a lot about myself and indirectly helped me to stay on track.

And the final words go to express my gratitude to my dear husband Péter. You have been my rock. And my shoulder to cry on. And my sound of reason. And my ultimate source of joy in life. Thank you for being my partner in life.

ABOUT THE AUTHOR

Anet Aselmaa was born in a small town called Võru in Estonia 30th October 1986. After secondary education, she moved to the capital to study Business Information Technology in Tallinn University of Technology. These studies gave her a global understanding of Information Technology. During her studies, she was interested in interaction design patterns and developed a web-based pattern management system as part of the Bachelor's thesis, and a design method for pattern based interface design as part of the Master's thesis. She received B.Sc. degree in 2007 and M.Sc. in 2010.

In 2006, next to her studies, she started working as a software developer and interaction designer. She worked primarily in the field of web-based software solutions. In addition to working in her home country, she also worked in Budapest, Hungary and Göteborg, Sweden. In October 2012, she decided to change the direction of her professional career and started as a PhD researcher in the Faculty of Industrial Design Engineering, Delft University of Technology.

The learnings and experiences from her time as a PhD researchers in the academia she will put into practice as a User Experience Designer. In her free time, she enjoys cooking and perfecting recipes for her family cookbook.

Publications by the author

Aselmaa A, van Herk M, Song Y, Goossens RHM, Laprie A (2017) The influence of automation on tumor contouring. *Cognition, Technology & Work*. doi: 10.1007/s10111-017-0436-0

Aselmaa A, van Herk M, Laprie A, Nestle U, Götz I, Wiedenmann N, Schimek-Jasch T, Picaud F, Syrykh C, Cagetti L V, Jolnerovski M, Song Y, Goossens RHM (2017) Using a contextualized sensemaking model for interaction design: A case study of tumor contouring. *Journal of Biomedical Informatics* 65:145–158 . doi: 10.1016/j.jbi.2016.12.001

Aselmaa A, Song Y, Goossens RHM (2016) User Interaction Visualization for Design Synthesis. *In: Isenberg T, Sadlo F (eds) Poster Proceedings of the*

Eurographics Conference on Visualization (EuroVis 2016). The Eurographics Association, p 3 doi: 10.2312/eurp.20161128

Aselmaa A, Goossens RHM, Rowland B, Laprie A, Song Y, Freudenthal A (2014) Medical Factors of Brain Tumor Delineation in Radiotherapy for Software Design. *In: Ahram T, Karwowski W, Marek T (eds) 5th International Conference on Applied Human Factors and Ergonomics (AHFE)*. pp 4865–4875, ISBN: 978-1-4951-2093-0

Aselmaa A, Song Y, Goossens RHM (2014) Design challenges in incorporating segmentation methods into radiotherapy software. *In: Multimodal imaging towards individualized radiotherapy treatments*. p 5, ISBN: 9789461863096

Ramkumar A, Aselmaa A, Song Y (2014) User interface concept and new workflow report. Delft University of Technology

Aselmaa A, Goossens RHM, Laprie A, Ramkumar A, Ken S, Freudenthal A (2013) External radiotherapy treatment planning–situation today and perspectives for tomorrow. *In: MASSOPTIER L, VIARD R (eds) Innovative imaging to improve radiotherapy treatments, 1st edn*. Lulu Enterprises, pp 91–98, ISBN: 9781291604177

Aselmaa A, Goossens RHM, Freudenthal A (2013) What is Sensemaking in the Context of External Radiotherapy Treatment Planning? *In: DMD Europe 2013: Design of Medical Devices Conference - Europe Edition*. Delft University of Technology, Delft, The Netherlands

Aselmaa A, Goossens RHM, Laprie A, Ken S, Fechter T, Ramkumar A, Freudenthal A (2013) Workflow Analysis Report. Delft University of Technology

Freudenthal, Adinda, Aselmaa Anet, Ramkumar Anjana, Laprie Anne, Massoptier Laurent. 2013. "Human Factors Engineering for Radiotherapy planning software." *In Innovative Imaging to Improve Radiotherapy Treatments, edited by Laurent MASSOPTIER and Romain VIARD, 1sted.*, 129-134. Lulu Enterprises.