



# Using a contextualized sensemaking model for interaction design

**DOI:**  
[10.1016/j.jbi.2016.12.001](https://doi.org/10.1016/j.jbi.2016.12.001)

**Document Version**  
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**  
Aselmaa, A., van Herk, M., Laprie, A., Nestle, U., Götz, I., Wiedenmann, N., Schimek-Jasch, T., Picaud, F., Strykh, C., Cagetti, L. V., Jolnerovski, M., Song, Y., & Goossens, R. H. M. (2017). Using a contextualized sensemaking model for interaction design: A case study of tumor contouring. *Journal of Biomedical Informatics*. <https://doi.org/10.1016/j.jbi.2016.12.001>

**Published in:**  
Journal of Biomedical Informatics

**Citing this paper**  
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**  
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**  
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



## Accepted Manuscript

Using a contextualized sensemaking model for interaction design: A case study of tumor contouring

Anet Aselmaa, Marcel van Herk, Anne Laprie, Ursula Nestle, Irina Götz, Nicole Wiedenmann, Tanja Schimek-Jasch, Francois Picaud, Charlotte Syrykh, Leonel V. Cagetti, Maria Jolnerovski, Yu Song, Richard H.M. Goossens

PII: S1532-0464(16)30175-7  
DOI: <http://dx.doi.org/10.1016/j.jbi.2016.12.001>  
Reference: YJBIN 2688

To appear in: *Journal of Biomedical Informatics*

Received Date: 5 May 2016  
Revised Date: 2 November 2016  
Accepted Date: 4 December 2016

Please cite this article as: 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, R.H.M., Using a contextualized sensemaking model for interaction design: A case study of tumor contouring, *Journal of Biomedical Informatics* (2016), doi: <http://dx.doi.org/10.1016/j.jbi.2016.12.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Using a contextualized sensemaking model for interaction design: a case study of tumor contouring

Anet Aselmaa <sup>a</sup>, Marcel van Herk <sup>b</sup>, Anne Laprie <sup>c</sup>, Ursula Nestle <sup>d</sup>, Irina Götz <sup>d</sup>, Nicole Wiedenmann <sup>d</sup>, Tanja Schimek-Jasch <sup>d</sup>, Francois Picaud <sup>c</sup>, Charlotte Syrykh <sup>c</sup>, Leonel V. Cagetti <sup>c</sup>, Maria Jolnerovski <sup>c</sup>, Yu Song <sup>a\*</sup>, Richard H.M. Goossens <sup>a</sup>

<sup>a</sup> Faculty of Industrial Design Engineering, Delft University of Technology  
Landbergstraat 15, 2628 CE Delft, The Netherlands

<sup>b</sup> The Netherlands Cancer Institute  
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

<sup>c</sup> Département de Radiothérapie, Institut Claudius-Regaud, Institut Universitaire du Cancer de Toulouse-Onco-pole,  
1, avenue Irène-Joliot-Curie, 31059 Toulouse, France

<sup>d</sup> Department for Radiation Oncology, University Medical Center Freiburg,  
Robert-Koch-Str. 3, 79106 Freiburg, Germany

\*corresponding author. E-mail: y.song@tudelft.nl

## Abstract

Sensemaking theories help designers understand the cognitive processes of a user when he/she performs a complicated task. This paper 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.

**Keywords:** interaction design; sensemaking; interaction patterns; radiotherapy; contouring

## Highlights

- We propose a two-step approach for incorporating sensemaking theory into HIS design.
- A contextual sensemaking (C-SM) model describing the tumor contouring was developed.
- Sensemaking and design insights regarding the tumor contouring task were generated.
- Five areas of improvements for increasing sensemaking support were identified.

## 1 Introduction

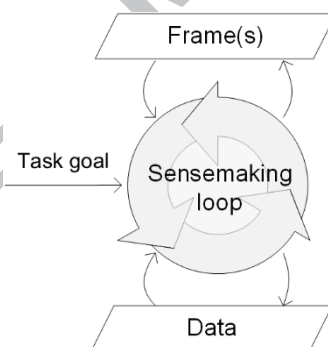
Health information systems (HIS) refer to computer based information systems (i.e., software and hardware) used in healthcare settings [1]. HIS were initially developed for patient care and administrative purposes, but are now being gradually extended to different areas of healthcare planning [2]. With 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 [3]. However, many of current solutions only offer limited support to typical cognitive tasks in the clinical domain, such as decision making, prevention of medical errors, etc. [4]

### 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 [5] 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” [6]. 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 [7]. Throughout a task, one is “facing gaps, building bridges across those gaps, evaluating outcomes and moving on” [8]. 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” [9].

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 Fig.1.



**Figure 1: 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 [10]: Human-Computer Interaction [7]; Cognitive Systems Engineering [9,11]; Organizational Communication [12] and Library and Information Science [8]. 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 [13] and CHI 2009 [14]). 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 [15].

The concept of information foraging, consisting of information seeking, gathering, and consumption [16], is closely associated to sensemaking. For instance, Pirolli and Card [17] 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. [18] 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 [7].

In the research area of applying sensemaking in the healthcare context, Mamykina et al. [19] 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 [20]. Besides, collaborative sensemaking had been applied in hospital emergency department setting [21], nursing [22], and online health forums [23]. Other specific areas of collaborative sensemaking that have been investigated are: team collaboration [24, 25], handoffs [26], 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.

## 1.2 Research approach

The aim of this paper 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 (Fig.2): 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.

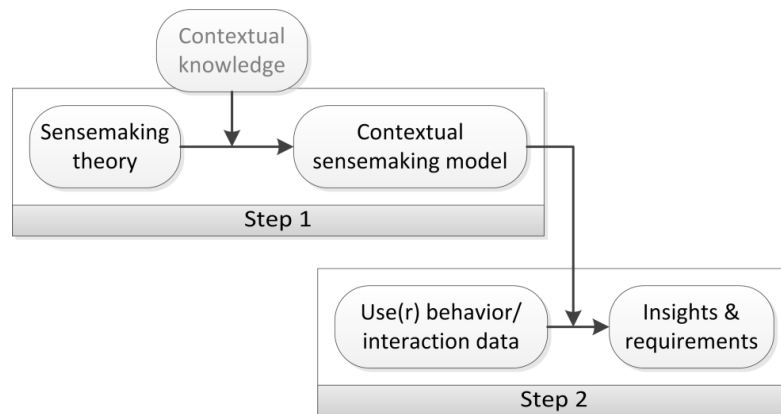


Figure 2: The proposed two-step approach

The remainder of this paper is structured as follows: In Section 2, based on observational research studies of the complicated tumor contouring task, the context of this task and the generalized sensemaking model from the literature, we developed the C-SM model. In Section 3, we present a case study where different navigation interactions and interaction sequence patterns were mapped to the developed C-SM model. The sensemaking and design insights obtained by incorporating the C-SM model into the analysis of navigation interactions and interaction sequence patterns are presented in Section 4. Finally, the outcomes and the proposed approach are discussed in Section 5.

## 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.

### 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 often navigate through these 2D images to construct the mental 3D model of the anatomy [27] 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 [28] 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. Within this workflow, one of the critical tasks that significantly influences the outcomes of the treatment is 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) [29]. 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 [30]. 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 [31], and large interobserver variabilities among physicians have been identified in several case studies (e.g., Fig.2). 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-Jaccard coefficient of the GTV volumes varied from 0.39 to 0.64 for nine cases [32]. This indicated a high interobserver variability among physicians, thus the final treatment plan highly depends on the judgement of individual physicians.



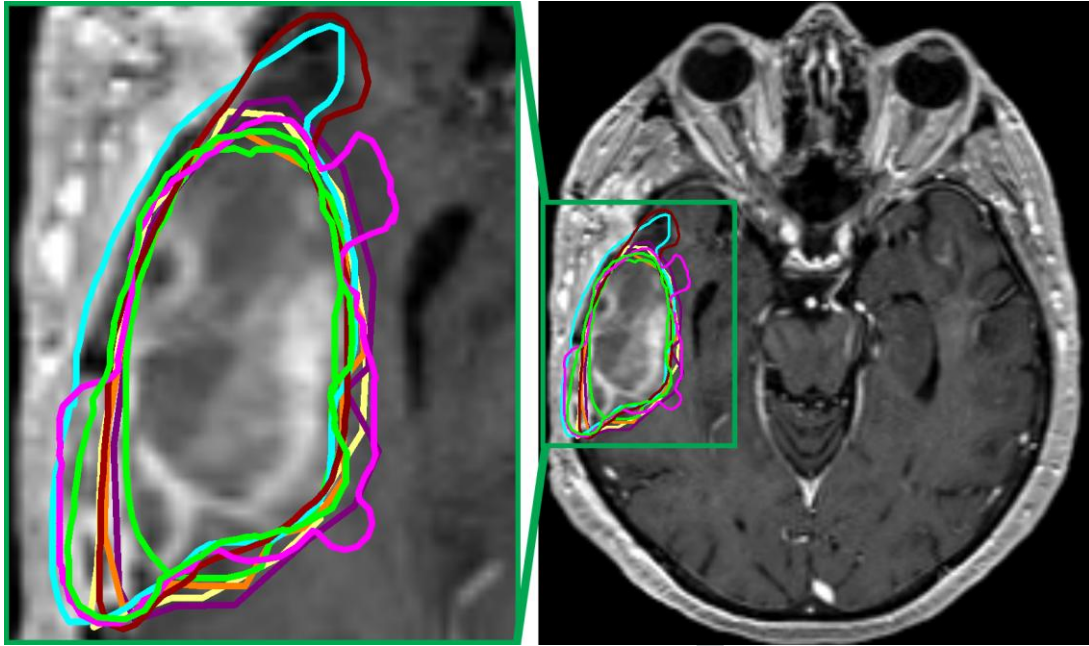


Figure 3: Example contours of a GTV on one 2D slice as contoured by eight physicians (each in different color) in a case of GBM, overlaid on MRI T1-weighted image with gadolinium enhancement. High interobserver variability can be observed.

The contouring task is cognitively demanding as there are multiple variables that the physicians need to take into consideration [33]. 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.

## 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 [34]. 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, respectively. 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 [28], three main task phases – named the familiarization phase, the action phase and the evaluation phase – were identified in the GTV contouring process as shown on Fig.4. 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.

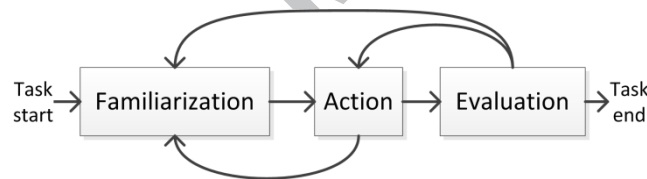


Figure 4: Three task phases in GTV contouring

The boundaries between 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 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 action) and then continues with evaluation.

### 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 Fig.1 was extended and adapted to the software's

use context first. Here, the model developed by Zhang et al. [5] 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 Fig.5. *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.

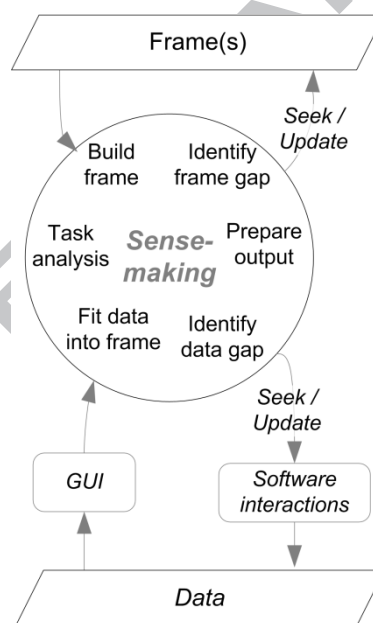


Figure 5: Generalized sensemaking model in the context of software use, including the sensemaking activities as identified by Zhang and Soergel [5].

Based on the identified GTV contouring task phases as illustrated in Fig.4 and the generalized sensemaking model in the context of software use as shown in Fig.5, 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 in Section 2.2. Figure 6 illustrates the resulting C-SM model.

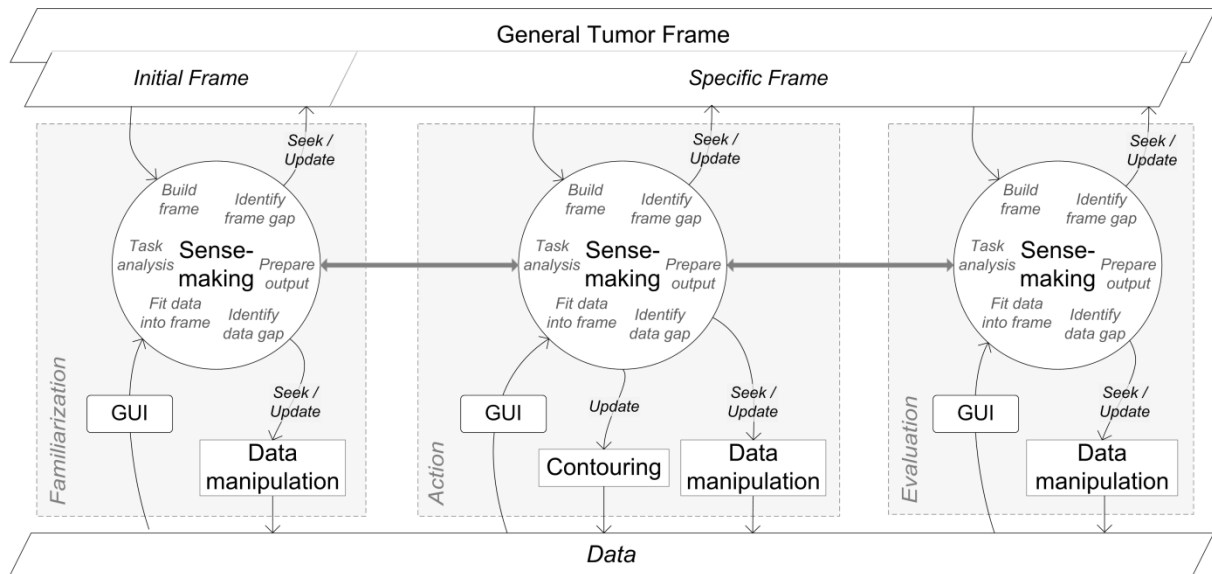


Figure 6: The contextualized sensemaking model (C-SM) of the tumor contouring task.

During tumor contouring, the frames involved represent primarily instances of a *general tumor frame* as the parallelogram at the top of Fig.6. 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.

### 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 as described in Section 2.1. 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. The outcomes of the analysis of the software interactions through the C-SM model are described in Section 4.

### 3.1 Materials and Methods

#### 3.1.1 The prototype

The case study was conducted with a software prototype (Fig.7) which was a modified and extended version of an existing contouring research software [35]. For each GTV contouring task, eight image datasets of a patient, which were in different modalities or acquired in different time during the treatment, were provided. They were: 1) pre-surgery MRI T1-weighted pre-gadolinium injection (MRI T1); 2) pre-surgery MRI T1-weighted post-gadolinium injection (MRI T1-gado); 3) pre-surgery MRI FLAIR; 4) radiotherapy treatment planning CT; 5) radiotherapy treatment planning MRI T1; 6) radiotherapy treatment planning MRI T1-gado; 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.

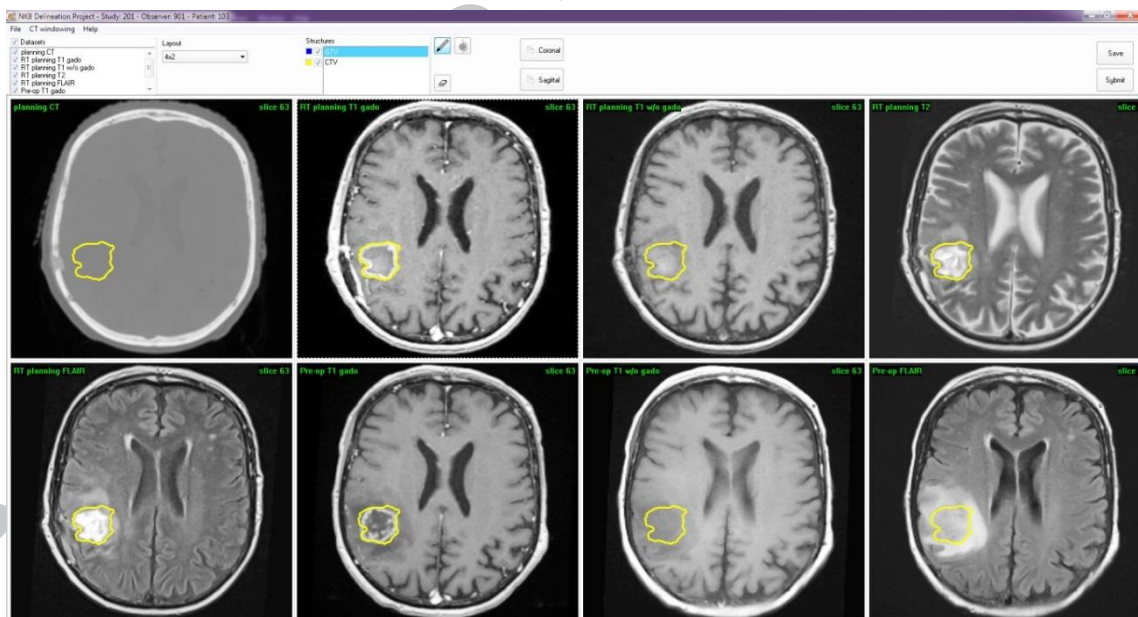


Figure 7: The GUI of the software prototype, with a layout representing the same 2D slice of all eight available datasets with a contour overlaid. Physicians typically used layout with 2-3 image datasets.

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 are 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 1). 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 dataset, 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 1: 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	Creation, modification or deletion of a contour
	Changing drawing tool	Switching from one drawing mode to another
	Interpolating	Generating a contour automatically based on neighboring contours

### 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-Onco-pole. 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” [36]. 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 a difficult case. Prior to the study tasks, the participants were given a training session in

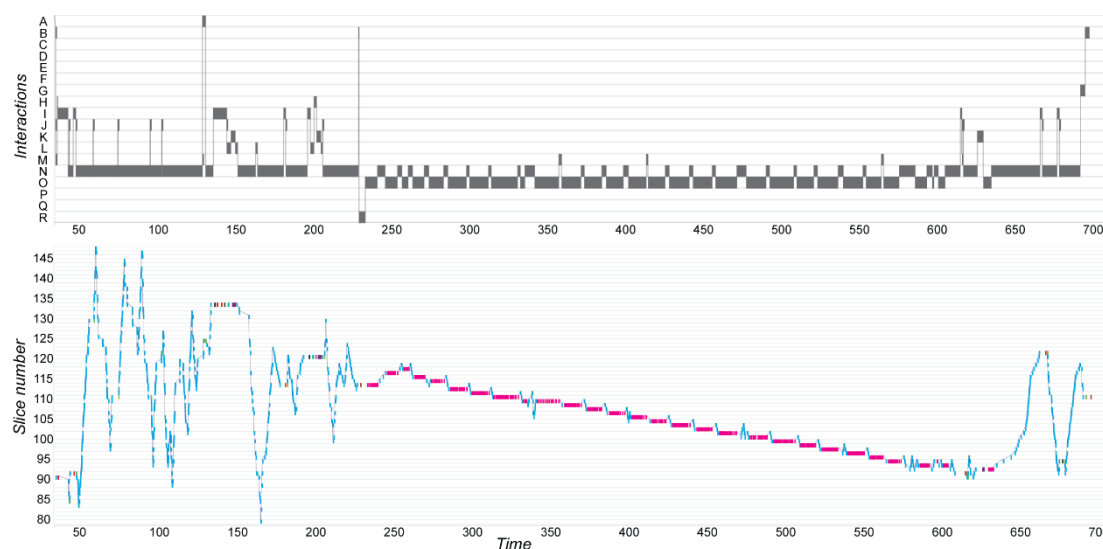
which they were also allowed to freely explore the software on another sample dataset. Ethical approval for using patient data for research purposes was obtained prior to the study. All physicians participating in the study were informed about the details of the study and signed informed consent forms 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-inch 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 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 of the software use was provided under the requests of physicians.

### 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 [37]. To enable this, a visual interaction log exploration tool was developed based on JavaScript and D3.js (<http://d3js.org>) [38]. 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 Fig.8. 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 8: 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 Fig.9. 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.

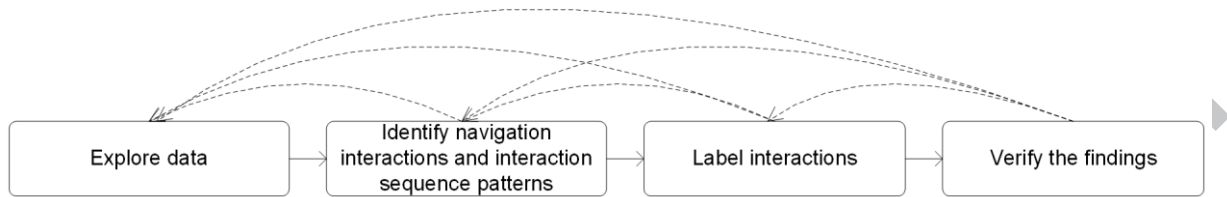


Figure 9: The process of identifying types of navigation interactions and interaction sequence patterns

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.

### 3.2 Results

The average task completion time was 11 minutes 26 seconds (SD = 6' 00"). Among the total task completion time, the average duration of the familiarization phase was 2 minutes 6 seconds (SD = 51"). The average duration of the action phase, which was calculated as the sum of contouring interactions, was 5 minutes 47 seconds (SD = 3' 47"). The rest of the time, on average 3 minutes 33

seconds (SD = 2' 00"), 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 (Fig.8) 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' kappa between two researchers was 0.957,  $p < 0.001$ ) and interaction sequence patterns occurred 388 times (the Cohen' kappa between two researchers was 0.785,  $p < 0.001$ ). Regarding each of the six cases, the Cohen' 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 Section 3.2.1 and Section 3.2.2, details of those identified navigation interactions and interaction sequence patterns will be presented, respectively.

### 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 milliseconds (ms, Standard Deviation (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 physician's 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.

**Table 2: The identified four types of navigation interactions t.p.s = time per slice, sec = second, ms = millisecond**

Types of Navigation interactions	Description	Task phase	Total count	Mean duration (sec)	Mean visible t.p.s (ms)	Mean slice change count
Single slice navigation	Scrolling that involved only one slice change	Familiarization	31	1.195	1195	1
		Evaluation	330	1.212	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

On average, the *single slice navigation* lasted for 1.211 seconds, the *neighbor navigation* for 2.8 seconds, the *region navigation* for 5.4 seconds, and the *long navigation* for 12.5 seconds. 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: 1.195" vs 1.212", 1.7" vs 2.8", 4.2" vs 5.6", and 11.2" vs 13.7"; 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 2). The *long navigation* represented rapid navigation through the datasets, during which one 2D image slice was visible on average 394 milliseconds. 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 milliseconds. 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 milliseconds 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.

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”.

### 3.2.2 Interaction sequence patterns

Through the visual analysis of the interaction logs, five interaction sequence patterns were identified as listed in Table 3.

**Table 3: Overview of the identified interaction sequence patterns.**

<b>Interaction sequence pattern</b>	<b>Interactions involved</b>	<b>Description</b>
Continuous zooming and panning	Zooming; Panning	The physician iteratively zooms and pans the datasets
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
Scrolling on a new dataset	Changing active dataset; Scrolling	The physician switches the dataset that they are scrolling through
Systematic contouring	Drawing; Changing drawing tool; Interpolating; Slice change; Scrolling	The physician is continuously drawing on neighboring slices
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

Table 4 presents the descriptive statistics of each of the patterns 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).

**Table 4: Descriptive statistics of the identified interaction sequence patterns; CE/PE ratio = the ratio of the duration of cognitive events to duration on of the physical events**

Pattern	Task phase	Total count	Mean duration (sec)	Mean t.p.s. (ms)	Mean slice change count	Mean CE/PE ratio
Continuous zooming and panning	Familiarization	19	8.9	-	-	1.1
	Evaluation	9	9.8	-	-	0.9
Data layout change before active dataset change	Familiarization	103	5.5	-	-	-
	Evaluation	36	2.8	-	-	-
Scrolling on a new dataset	Familiarization	148	8.4	405	27.6	10.9
	Evaluation	82	7.3	490	18.9	10.0
Systematic contouring	Evaluation-Action	242	33.3	4356	6.8	1.3
Scrolling which results in a single slice contouring	Evaluation-Action	133	14.4	-	13.3	3.2

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 seconds. 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 Fig.10. 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.

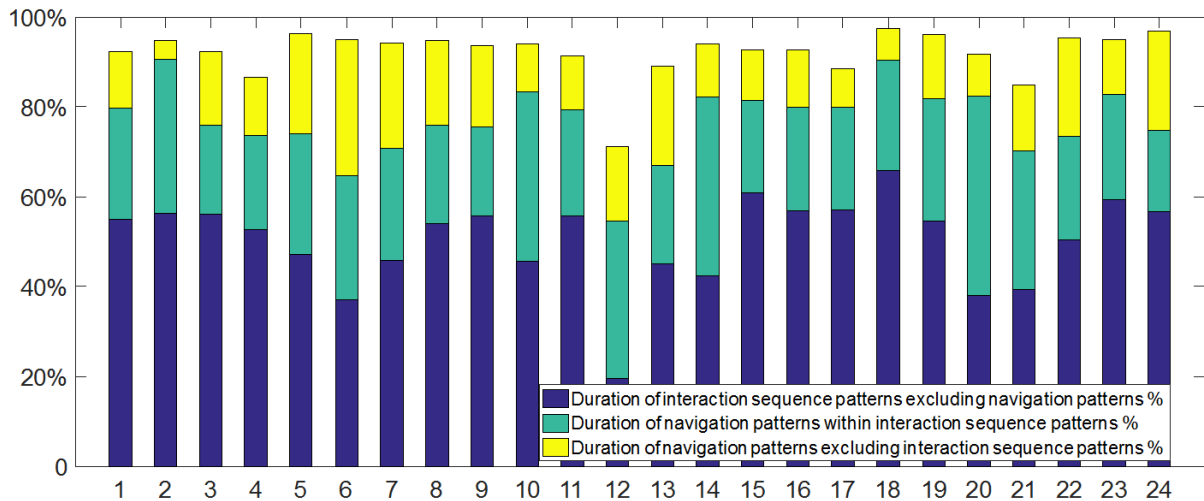


Figure 10: Total coverage by the navigation interactions and the interaction sequence patterns in relation to the task completion time (100%) for all 24 cases.

#### 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 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 (Section 4.1) and to generate requirements for the interaction design (Section 4.2).

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

Type of navigation interactions	Task phase	Inferred sensemaking activity	Indication of the sensemaking activity	Design insight (category)
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)
<b>Type of interaction sequence patterns</b>				
Continuous zooming and panning	Familiarization	Focused information seeking	Increased magnification level. When the magnification level increases, one's viewing is more focused [39]	Reduce time and physical effort (5) Support detecting regions of interest (3)
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 ( <i>contouring interactions</i> )	Action	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	Focused information (gap) seeking	Navigating within the proximal data	Support contour evaluation in 3D space (4)
Scrolling which results in a single slice contouring ( <i>contouring interactions</i> )	Action	Preparing the output	Updating existing data	Support identifying regions for correction (3)
Scrolling which results in a single slice contouring ( <i>navigation interaction</i> )	Evaluation	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

## 4.1 Sensemaking insights

In this section, we attempt to bring connections among the sensemaking activities (as shown in Fig.6), 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.

### 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 influences 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 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's 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.

#### 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 phase, the transitions between the action phase and the evaluation phase were fuzzy and more frequent. As a result, physicians had typically more than one contouring episode (i.e., continuous contouring interactions).

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 [40]. 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*).

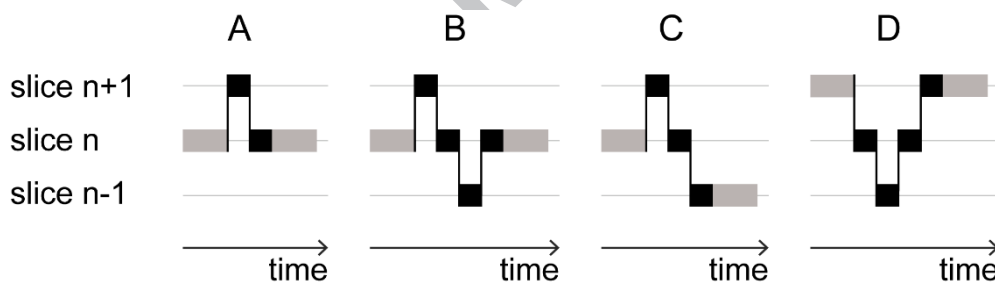
Depending on the personal preferences, the specific contouring intention and the available data, the physicians engaged in different contouring strategies (result from the *task analysis* activity). All physicians were engaged in *systematic contouring* to some extent. Some physicians took a “precise” 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 Fig.8). 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.

### 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 Fig.11.



**Figure 11: 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.

## 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.

### 4.2.1 Using the C-SM model to generate design insights

The main focus of using 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 proposed to use the CE/PE ratio (as seen in Table 4) 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.

### 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 6 highlights the primary indicators for sensemaking support improvements and their types. The main design requirements for supporting sensemaking can be categorized into 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, corresponding to the sensemaking activities which they support, respectively.

**Table 6: Overview of potential indicators for improvements regarding effectiveness and efficiency.**

Indicator for possible improvements	Types of improvements
<ul style="list-style-type: none"> <li>• Long duration of cognitive events</li> <li>• High frequency of cognitive events</li> <li>• High cognitive involvement</li> <li>• Use of external tools/materials</li> </ul>	Effectiveness
<ul style="list-style-type: none"> <li>• Long duration</li> <li>• Alternating cognitive events with physical events</li> </ul>	Effectiveness and Efficiency
<ul style="list-style-type: none"> <li>• Long overall duration</li> <li>• Low cognitive involvement (short durations of cognitive events)</li> <li>• Repetition of the same type of interaction</li> </ul>	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, 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 Discussion

### 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 [37]. 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 [41]. However, we have shown that identifying patterns is beneficial for generating in-depth 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 [42]. Furthermore, most of those algorithms still require human involvement [43]. 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 do 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 [44].

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 [45] 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 The two-step approach

We proposed a two-step approach for incorporating sensemaking in order to identify additional design inputs. The first step is to model the sensemaking in a context, where a 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. Through considerable time and efforts were spent on the research, it offered rich contextual information and made the task tangible to 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) [46]. 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.

### 5.3 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 [47].

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.

### 5.4 Conclusion

In this paper, 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.

## Acknowledgement

The presented research is part of the Software for the Use of Multi-Modality images in External Radiotherapy (SUMMER) project which is funded by European Commission (FP7-PEOPLE-2011-ITN) under grant agreement PITN-GA-2011-290148.

The authors would also like to express their appreciations to Mr. Joop C. Duppen, who developed the original “Big Brother” software, and Mr. Lennert Ploeger, for technical support during software prototype development. The authors would also like to thank Mr. Peter Kun, for his constructive comments during the preparation of the manuscript.

## References

- [1] M.M. Yusof, A. Papazafeiropoulou, R.J. Paul, L.K. Stergioulas, Investigating evaluation frameworks for health information systems, *Int. J. Med. Inform.* 77 (2008) 377–385. doi:<http://dx.doi.org/10.1016/j.ijmedinf.2007.08.004>.
- [2] R. Haux, Health information systems – past, present, future, *Int. J. Med. Inform.* 75 (2006) 268–281. doi:<http://dx.doi.org/10.1016/j.ijmedinf.2005.08.002>.
- [3] V.L. Patel, T.G. Kannampallil, Cognitive informatics in biomedicine and healthcare, *J. Biomed. Inform.* 53 (2015) 3–14. doi:<http://dx.doi.org/10.1016/j.jbi.2014.12.007>.



- [4] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, I. Winblad, National questionnaire study on clinical ICT systems proofs: Physicians suffer from poor usability, *Int. J. Med. Inform.* 80 (2011) 708–725. doi:<http://dx.doi.org/10.1016/j.ijmedinf.2011.06.010>.
- [5] P. Zhang, D. Soergel, Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking, *J. Assoc. Inf. Sci. Technol.* 65 (2014) 1733–1756. doi:<http://dx.doi.org/10.1002/asi.23125>.
- [6] Y. Malhotra, Expert systems for knowledge management: crossing the chasm between information processing and sense making, *Expert Syst. Appl.* 20 (2001) 7–16. doi:[http://dx.doi.org/10.1016/S0957-4174\(00\)00045-2](http://dx.doi.org/10.1016/S0957-4174(00)00045-2).
- [7] D.M. Russell, M.J. Stefik, P. Pirolli, S.K. Card, The cost structure of sensemaking, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '93*, ACM Press, New York, New York, USA, 1993: pp. 269–276. doi:<http://dx.doi.org/10.1145/169059.169209>.
- [8] B. Dervin, Sense-making theory and practice: an overview of user interests in knowledge seeking and use, *J. Knowl. Manag.* 2 (1998) 36–46. doi:<http://dx.doi.org/10.1108/13673279810249369>.
- [9] G. Klein, B. Moon, R.R. Hoffman, Making sense of sensemaking II: A macrocognitive model., *IEEE Intell. Syst.* 21 (2006) 88–92. doi:<http://dx.doi.org/10.1109/MIS.2006.100>.
- [10] B. Dervin, C.M. Naumer, Sense-Making, in: *Encycl. Libr. Inf. Sci.*, Third Edit, Taylor & Francis, 2009: pp. 4696–4707. <http://www.tandfonline.com/doi/abs/10.1081/E-ELIS3-120043227>.
- [11] G. Klein, J.K. Phillips, E.L. Rall, D.A. Peluso, A data-frame theory of sensemaking, in: R.R. Hoffman (Ed.), *Expert. out Context Proc. Sixth Int. Conf. Nat. Decis. Mak.*, Psychology Press, 2007: pp. 113–155.
- [12] K.E. Weick, *Sensemaking in organizations*, Sage, 1995.
- [13] D. Russell, Sensemaking Workshop CHI 2008 Extended Abstract, in: *Sensemaking Work.*, ACM Press, New York, New York, USA, 2008: pp. 3981–3984. doi:<http://dx.doi.org/10.1145/1358628.1358972>.
- [14] D.M. Russell, P. Pirolli, G. Furnas, S.K. Card, M. Stefik, Sensemaking workshop CHI 2009, in: *Proc. 27th Int. Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '09*, ACM Press, Boston, MA, 2009: p. 4751. doi:<http://dx.doi.org/10.1145/1520340.1520732>.
- [15] P. Pirolli, D. Russell, Introduction to this Special Issue on Sensemaking, *Human-Computer Interact.* 26 (2011) 1–8. doi:<http://dx.doi.org/10.1080/07370024.2011.556557>.
- [16] P. Pirolli, S. Card, Information Foraging, *Psychological Rev.* 106 (1999) 643–675. doi:<http://dx.doi.org/10.1037/0033-295x.106.4.643>.
- [17] P. Pirolli, S. Card, The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis, in: *Proc. Int. Conf. Intell. Anal.*, 2005: pp. 2–4.
- [18] T.G. Kannampallil, A. Franklin, R. Mishra, K.F. Almoosa, T. Cohen, V.L. Patel, Understanding the nature of information seeking behavior in critical care: Implications for the design of health information technology, *Artif. Intell. Med.* 57 (2013) 21–29. doi:<http://dx.doi.org/10.1016/j.artmed.2012.10.002>.
- [19] L. Mamykina, A.M. Smaldone, S.R. Bakken, Adopting the sensemaking perspective for chronic

- disease self-management, *J. Biomed. Inform.* 56 (2015) 406–417. doi:<http://dx.doi.org/10.1016/j.jbi.2015.06.006>.
- [20] C. Bossen, L.G. Jensen, How physicians “achieve overview”: A case-based study in a hospital ward, *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW.* (2014) 257–268. doi:<http://dx.doi.org/10.1145/2531602.2531620>.
- [21] S.A. Paul, M.C. Reddy, *Understanding Together : Sensemaking in Collaborative Information Seeking, Communication. The Pennsy* (2010) 321–330. doi:<http://dx.doi.org/10.1145/1718918.1718976>.
- [22] M. Kristiansen, A. Obstfelder, A.T. Lotherington, Nurses’ sensemaking of contradicting logics: An underexplored aspect of organisational work in nursing homes, *Scand. J. Manag.* 31 (2015) 330–337. doi:<http://dx.doi.org/10.1016/j.scaman.2015.04.003>.
- [23] L. Mamykina, D. Nakikj, N. Elhadad, Collective Sensemaking in Online Health Forums, in: *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI ’15*, ACM Press, New York, New York, USA, 2015: pp. 3217–3226. doi:<http://dx.doi.org/10.1145/2702123.2702566>.
- [24] M.C. Reddy, B.J. Jansen, A model for understanding collaborative information behavior in context: A study of two healthcare teams, *Inf. Process. Manag.* 44 (2008) 256–273. doi:<http://dx.doi.org/10.1016/j.ipm.2006.12.010>.
- [25] L.K. Leykum, H. Chesser, H.J. Lanham, P. Carla, R. Palmer, T. Ratcliffe, H. Reisinger, M. Agar, J. Pugh, The Association Between Sensemaking During Physician Team Rounds and Hospitalized Patients??? Outcomes, *J. Gen. Intern. Med.* 30 (2015) 1821–1827. doi:<http://dx.doi.org/10.1007/s11606-015-3377-4>.
- [26] N. Sharma, Sensemaking handoff: When and how?, *Proc. Am. Soc. Inf. Sci. Technol.* 45 (2009) 1–12. doi:<http://dx.doi.org/10.1002/meet.2008.1450450234>.
- [27] E. Varga, P.M.T. Pattynama, A. Freudenthal, Manipulation of mental models of anatomy in interventional radiology and its consequences for design of human–computer interaction, *Cogn. Technol. Work.* 15 (2013) 457–473. doi:<http://dx.doi.org/10.1007/s10111-012-0227-6>.
- [28] A. Aselmaa, R.H.M. Goossens, A. Laprie, S. Ken, T. Fechter, A. Ramkumar, A. Freudenthal, *Workflow Analysis Report*, Delft University of Technology, 2013. <http://summer-project.eu/work/deliverables/>.
- [29] N.G. Burnet, Defining the tumour and target volumes for radiotherapy, *Cancer Imaging.* 4 (2004) 153–161. doi:<http://dx.doi.org/10.1102/1470-7330.2004.0054>.
- [30] C. Nutting, D.P. Dearnaley, S. Webb, Intensity modulated radiation therapy: a clinical review., *Br. J. Radiol.* 73 (2000) 459–469. doi:<http://dx.doi.org/10.1259/bjr.73.869.10884741>.
- [31] C.F. Njeh, Tumor delineation: The weakest link in the search for accuracy in radiotherapy, *J. Med. Phys.* 33 (2008) 136–140. <http://www.jmp.org.in/text.asp?2008/33/4/136/44472>.
- [32] C.W. Wee, W. Sung, H.-C. Kang, K.H. Cho, T.J. Han, B.-K. Jeong, J.-U. Jeong, H. Kim, I.A. Kim, J.H. Kim, S.H. Kim, S. Kim, D.S. Lee, M.Y. Lee, D.H. Lim, H.L. Park, C.-O. Suh, S.M. Yoon, I.H. Kim, Evaluation of variability in target volume delineation for newly diagnosed glioblastoma: a multi-institutional study from the Korean Radiation Oncology Group, *Radiat. Oncol.* 10 (2016) 137. doi:<http://dx.doi.org/10.1186/s13014-015-0439-z>.
- [33] A. Aselmaa, R.H.M. Goossens, B. Rowland, A. Laprie, Y. Song, A. Freudenthal, *Medical Factors*

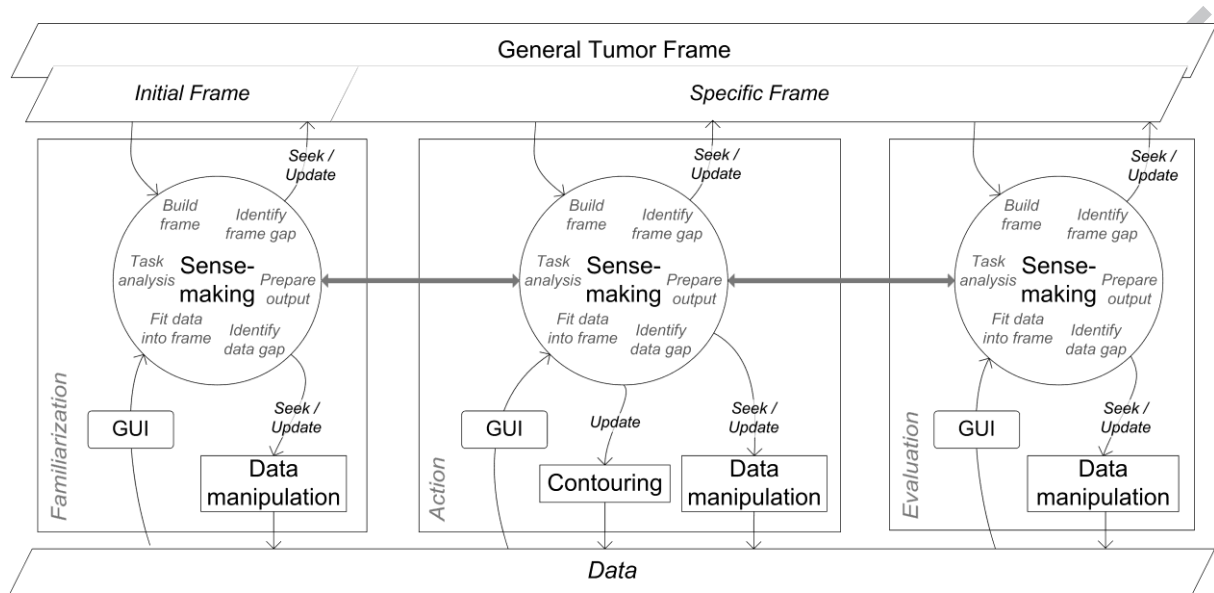
- of Brain Tumor Delineation in Radiotherapy for Software Design, in: T. Ahram, W. Karwowski, T. Marek (Eds.), 5th Int. Conf. Appl. Hum. Factors Ergon., 2014: pp. 4865–4875.
- [34] S. Kim, D. Soergel, Selecting and measuring task characteristics as independent variables, *Proc. Am. Soc. Inf. Sci. Technol.* 42 (2006) n/a-n/a. doi:<http://dx.doi.org/10.1002/meet.14504201111>.
- [35] R.J.H.M. Steenbakkers, J.C. Duppen, I. Fitton, K.E.I. Deurloo, L. Zijp, A.L.J. Uitterhoeve, P.T.R. Rodrigus, G.W.P. Kramer, J. Bussink, K. De Jaeger, J.S.A. Belderbos, A.A.M. Hart, P.J.C.M. Nowak, M. van Herk, C.R.N. Rasch, Observer variation in target volume delineation of lung cancer related to radiation oncologist–computer interaction: A “Big Brother” evaluation, *Radiother. Oncol.* 77 (2005) 182–190. doi:<http://dx.doi.org/10.1016/j.radonc.2005.09.017>.
- [36] M. Niyazi, M. Brada, A.J. Chalmers, S.E. Combs, S.C. Erridge, A. Fiorentino, A.L. Grosu, F.J. Lagerwaard, G. Minniti, R.-O. Mirimanoff, U. Ricardi, S.C. Short, D.C. Weber, C. Belka, ESTRO-ACROP guideline “target delineation of glioblastomas,” *Radiother. Oncol.* (2016). doi:<http://dx.doi.org/10.1016/j.radonc.2015.12.003>.
- [37] W. Dou, D.H. Jeong, F. Stukes, W. Ribarsky, H.R. Lipford, R. Chang, Recovering Reasoning Processes from User Interactions, *IEEE Comput. Graph. Appl.* 29 (2009) 52–61. doi:<http://dx.doi.org/10.1109/MCG.2009.49>.
- [38] A. Aselmaa, Y. Song, R.H.M. Goossens, User Interaction Visualization for Design Synthesis, in: T. Isenberg, F. Sadlo (Eds.), *Poster Proc. Eurographics Conf. Vis. (EuroVis 2016)*, The Eurographics Association, 2016: p. 3. doi:<http://dx.doi.org/10.2312/eurp.20161128>.
- [39] L. Pantanowitz, A. Parwani, E. Tseytlin, C. Mello-Thoms, Analysis of slide exploration strategy of cytologists when reading digital slides, in: C.K. Abbey, C.R. Mello-Thoms (Eds.), *International Society for Optics and Photonics*, 2012: p. 831800. doi:10.1117/12.912498.
- [40] S. Zabramski, Careless touch, in: *Proc. 23rd Aust. Comput. Interact. Conf. - OzCHI '11*, ACM Press, New York, New York, USA, 2011: pp. 329–332. doi:<http://dx.doi.org/10.1145/2071536.2071588>.
- [41] W. Jorritsma, F. Cnossen, R.A. Dierckx, M. Oudkerk, P.M.A. van Ooijen, Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client, *Int. J. Med. Inform.* 85 (2016) 36–42. doi:<http://dx.doi.org/10.1016/j.ijmedinf.2015.10.007>.
- [42] S. Bauer, R. Wiest, L.-P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, *Phys. Med. Biol.* 58 (2013) R97–R129. doi:[dx.doi.org/10.1088/0031-9155/58/13/R97](http://dx.doi.org/10.1088/0031-9155/58/13/R97).
- [43] F. Heckel, J.H. Moltz, C. Tietjen, H.K. Hahn, Sketch-Based Editing Tools for Tumour Segmentation in 3D Medical Images, *Comput. Graph. Forum.* 32 (2013) 144–157. doi:<http://dx.doi.org/10.1111/cgf.12193>.
- [44] I. Fitton, S.A.P. Cornelissen, J.C. Duppen, R.J.H.M. Steenbakkers, S.T.H. Peeters, F.J.P. Hoebbers, J.H.A.M. Kaanders, P.J.C.M. Nowak, C.R.N. Rasch, M. van Herk, Semi-automatic delineation using weighted CT-MRI registered images for radiotherapy of nasopharyngeal cancer, *Med Phys.* 38 (2011) 4662–4666. doi:<http://dx.doi.org/10.1118/1.3611045>.
- [45] F. Zhao, X. Xie, An Overview of Interactive Medical Image Segmentation, *Ann. BMVA.* 2013 (2013) 1–22.
- [46] P.H. Lindsay, D.A. Norman, 14 - Problem solving, in: P.H. Lindsay, D.A. Norman (Eds.), *Hum.*

Inf. Process., Academic Press, 1972: pp. 500–525. doi:<http://dx.doi.org/10.1016/B978-0-12-450950-4.50019-6>.

- [47] M. van den Haak, M. De Jong, P. Jan Schellens, Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue, *Behav. Inf. Technol.* 22 (2003) 339–351. doi:<http://dx.doi.org/10.1080/0044929031000>.

ACCEPTED MANUSCRIPT

## Graphical abstract



## Highlights

- We propose a two-step approach for incorporating sensemaking theory into HIS design.
- A contextual sensemaking (C-SM) model describing the tumor contouring was developed.
- Sensemaking and design insights regarding the tumor contouring task were generated.
- Five areas of improvements for increasing sensemaking support were identified.