Enhancing Accuracy, Time Spent, and Ubiquity in Critical Healthcare Delineation via Cross-Device Contouring

Matin Yarmand
UCSD The Design Lab
La Jolla, CA, USA
myarmand@ucsd.edu

Chen Chen
UCSD The Design Lab
La Jolla, CA, USA
chenchen@ucsd.edu

Michael V. Sherer
UCSD Radiation Medicine
La Jolla, CA, USA
msherer@health.ucsd.edu

Yash N. Shah
UCSD Computer Science
La Jolla, CA, USA
ynshah@ucsd.edu

Peter Liu
UCSD Computer Science
La Jolla, CA, USA
p3liu@ucsd.edu

Borui Wang
UW HCDE
Seattle, WA, USA
theworld@uw.edu

Larry Hernandez
UCSD Radiation Medicine
La Jolla, CA, USA
larryhernandez@uwalumni.com

James D. Murphy
UCSD Radiation Medicine
La Jolla, CA, USA
j2murphy@health.ucsd.edu

Nadir Weibel
UCSD The Design Lab
La Jolla, CA, USA
weibel@ucsd.edu

ABSTRACT

Improving accuracy, time spent, and ubiquity of delineation has been a long-standing design aim, yet many HCI works have overlooked high-stakes and complex healthcare annotation. We explore *contouring*, a critical workflow aimed at identifying and segmenting tumors, usually performed on immobile desktop computers in clinics, in which limited support for mobile access leads to prolonged and subpar treatment planning. Following interviews and think-aloud studies (N = 10 physicians), we report key contouring behaviors, and later design a novel cross-device prototype that enables contouring on everyday touch devices. We compared contouring via desktop and touch in a lab study (N = 8 residents) and found that mobile phones not only yielded similar accuracy, but also took significantly less time. Our results point to three broad design guidelines for cross-device solutions deployed within standalone healthcare workflows, and highlight how incorporating different device and input modalities can improve treatment delivery in today's distributed healthcare environments.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**.

KEYWORDS

Healthcare, Contour Delineation, Cross-Device Interface Design

ACM Reference Format:
Matin Yarmand, Chen Chen, Michael V. Sherer, Yash N. Shah, Peter Liu, Borui Wang, Larry Hernandez, James D. Murphy, and Nadir Weibel. 2024. Enhancing Accuracy, Time Spent, and Ubiquity in Critical Healthcare Delineation via Cross-Device Contouring. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3643834.3660718

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3660718

1 INTRODUCTION

Exploring the trade-offs between accuracy, time spent, and mobile access (or ubiquity) — the three determining factors of effective delineation — has been an important line of research in human-computer interaction and design of interactive systems which often exhibit non-conforming patterns in different contexts. Many prior studies that aimed to evaluate delineation performance of input modalities (*i.e.,* mouse, finger-touch, and stylus-touch) pointed out that while touch-based interactions are typically faster, they can be more error prone [33, 61, 62]. Also, while stylus-based sketching can be more accurate than finger-touch interactions [63], the need for an external device (*i.e.,* stylus) can introduce additional cost and complexities. As mobile computing aims to maximize the level of mobility [40], integrating additional hardware can also degrade the capability to physically move and utilize computing services. In addition, the type of task (*e.g.,* precision-first drawing [19] *vs.* sketching), as well as user context (*e.g.,* age and experience level) can further impact design decisions that uniquely support the participating user group [27, 34]. As such, understanding the full context of the drawing task, as well as the common practices and tendencies, is a crucial step in designing delineation interfaces that address and optimize the existing trade-offs among these three factors.

Beyond general drawing applications (explored by large in prior works), delineation is an indispensable component for many high-stakes healthcare tasks and workflows, yet have received little attention in prior work. One such task is *contouring*, which refers to the process whereby an oncologist identifies and outlines tumors and normal organs from a stack of medical images, producing personalized radiation treatment plans for cancer patients. Safe and effective radiation therapy depends critically on high quality contouring. The default computerized support for contouring is standalone and immobile desktop with mouse devices which are mainly located in clinics.

Contouring involves a complex mix of accuracy, time spent, and needs for ubiquitous access which directly contributes to treatment delivery and patient well-being. *First*, treatment plans define the foundation for delivering high-energy radiation to patients and require sub-millimeter precision. Inaccurate radiation planning results in detrimental consequences for patient survival [48, 56]. The main challenge is maximizing radiation to tumor regions while minimizing the impact on the surrounding organs at risk, and accurate contouring is key to enabling such localized radiation. *Second*, contouring is already a time-intensive process, taking on average up to 30 hours per patient [7], and the limited access to contouring stations can exacerbate the existing delay in treatment planning [21]. *Third*, as patients' anatomy and physiology change during the treatment course [37], frequent re-generation of treatment plans, and hence re-contouring of the tumoral regions is needed to avoid a mismatch between the tumor being treated and the applied radiation. While recent AI-based automatic contouring approaches (*e.g.,* [36, 43]) can facilitate a more efficient contouring workflow, their lack of precision and limited support for real patient cases (with unique disease circumstances) constrain widespread adoption [58]. Given the existing infrastructure — and the high cost in time and availability of desktop-based contouring tools [38] — generating new contours and radiation plans on a daily basis is challenging.

To maintain high accuracy, lower time spent, and improve the ubiquity of existing contouring workflows, we introduce a cross-device approach to contouring, and assess its feasibility in this critical healthcare domain. We designed interfaces that robustly support contouring, not just on the default desktop and mouse, but also on everyday touch devices. This work more broadly explores the following Research Questions (RQs):

RQ1: *what are common contouring behaviors (e.g., action types), input devices, and environment set-ups?*

RQ2: *How can the design of software solutions effectively balance accuracy, time efficiency, and ubiquity?*

RQ3: *what general design guidelines can transform existing healthcare workflows via cross-device interfaces?*

Following semi-structured interviews and think-aloud studies with four faculty and six residents from the Department of Radiation Medicine at the UC San Diego Health, this paper defines three main categories of contouring actions and common patterns, reports promising empirical evidence for touch-based contouring despite the default method with desktop and mouse, and discusses trade-offs of environment set-ups with different image views and types (**RQ1**; Sec. 3.2). Guided by these findings, we design and develop a prototype for *cross-device* contouring, in which physicians can use their everyday touch devices to contour on a simplified, yet effective UI design (which encapsulates the main contouring actions) and leverage the existing patterns. A lab study (N = 8 residents) revealed that everyday touch devices are not only viable for contouring, but also they perform on-par or better than the traditional Desktop & Mouse in terms of time spent (30.6% less in Phone & Finger), accuracy (similar across all conditions), and system usability (28.5% better in Tablet & Stylus) (**RQ2**; Sec. 6). These results point to the potential of adopting a contouring environment that facilitates fluid and ubiquitous contouring across different configurations of device and input modalities. We further offer three design principles to integrate cross–device functionality in other static healthcare domains: convenience, consistency, and balance (**RQ3**; Sec. 7.3).

2 BACKGROUND AND RELATED WORK

This section provides background information on radiotherapy treatment and the existing contouring tools, as well as a summary of prior works that explored the impacts of input modalities on delineation tasks.

2.1 Contouring in Radiotherapy Treatment Planning

Radiotherapy is the medical practice of treating cancer by delivering high dose radiation to the tumor. The current procedure contains a multi-disciplinary workflow of multiple clinicians with specialized tasks [8]: broadly, once organ-specific physicians diagnose patients and detect cancer, a group of medical practitioners (including radiation oncologists and surgeons) discuss details of treatment and request the acquisition of specific image sets (*e.g.,* CT scans). Later, radiation oncologists delineate regions on the medical images (the process known as *contouring*) and distinguish between malignant tumors and the surrounding organs at risk [16]. Lastly, radiation therapists deliver the prescribed dose according to the treatment plans.

Radiation oncologists perform contouring — using only desktop based tools such as MIM [3] and Eclipse [2] — by drawing 2D contours repeatedly on relevant image slices to encompass the 3D volume of the tumorous tissues. As displayed in Figure 1, these tools facilitate defining structures, delineating regions, scrolling through slices, as well as pre-contouring (*e.g.,* selecting colors) and post-contouring (*e.g.,* applying fixed margin) customization. Few recent research has offered solutions to enhance contouring efficiency using virtual reality [17, 18] and AI-based medical image segmentation [49]. However, these solutions either require cumbersome hardware or lack adequate accuracy. Further, existing pre-trained AI models for automatic contouring can only provide a generalized contouring solution (*e.g.,* [36] for cervix and [43] for prostate) and fail to support unique patient cases (while considering granular patient history) with high precision. As such, manual contouring remains the status quo in real-world clinical practices, especially given that tumor identification contains substantial uncertainty, in which physicians have to rely on their unique experience and tendencies [58].

Contouring is a cognitively demanding procedure that calls for extreme precision, and clinicians need to take into account three categories of medical factors: treatment context (*e.g.,* clinical symptoms), tumor context (*e.g.,* size and growth direction of tumor), and tumorous areas (*e.g.,* satellite regions) [9]. The high cognitive demand of contouring, as well as the lack of flexible hardware and software support can lead to long contouring sessions and diminished accuracy. This paper investigates existing practices of radiation oncologists to uncover unique contouring

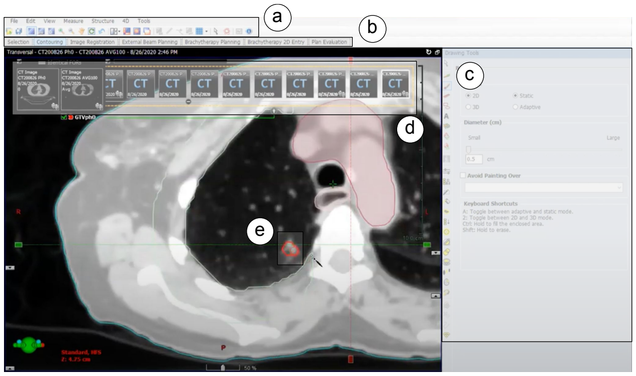


Figure 1: Eclipse interface, a common contouring software. Key interface components include (a) image manipulation and system settings, such as zooming in/out; (b) general workflow selection (*e.g.,* contouring and image fusion); (c) set of delineation tools including the commonly used pencil, eraser, and interpolation; (d) prior treatment plans of the same case; and (e) the drawn contour using the pencil tool. While this application contains many tools on the top and right sidebars, physicians rarely use many of these features. Compared to Eclipse which is only compatible with immobile desktop computers, we designed a cross–device contouring prototype that runs on any desktop and touch devices with a consistent interface that addresses the core contouring needs.

actions and patterns, and how these insights can lower cognitive load and improve contouring experience. We further show that well-designed *cross-device contouring* (according to unique needs of practitioners) provides flexibility via everyday touch devices and also facilitates fast and accurate delineation beyond just desktop interfaces.

2.2 Input-Based Drawing and Dragging Performance

Prior works have extensively measured the performance of input modalities (*i.e.,* mouse, touch, and stylus) for drawing and dragging tasks, and reported non-conforming completion time and error rates. Early work incorporated tracing sine waves which yielded shortest time and highest preference for mouse over stylus, and then touch [42]. Other works reported that using a mouse is generally faster than interacting with a touch device [28], but also show how accuracy is often more stable across touch devices [20]. When exploring the impact of touch, stylus, and mouse on drawing performance, Zabramski and Stuerzlinger [63] revealed how using a pen is consistently a better input modality, while touch is more error-prone. Other research incorporated more robust drawing tasks (*i.e.,* tracing over randomly generated shapes) and pointed to the speed-accuracy trade-off: touch-based interaction is faster, yet less accurate [61, 62]. Mouse-based drawing, however, produced significantly higher errors in curved shapes compared to linear tasks [33].

As evident from the varying findings in the existing works, identifying the optimal device and input modality highly depends

Table 1: Background details on the four faculty and six residents who participated in this study.

ID	Title	Gender	Age	Experience
F1	Assistant Professor	Male	32	5 years
F2	Assistant Professor	Female	39	11 years
F3	Assistant Professor	Male	34	6 years
F4	Assistant Professor	Male	37	7 years
R1	Resident Year 4	Male	35	4 years
R2	Resident Year 4	Female	33	4 years
R3	Resident Year 1	Female	28	6 months
R4	Resident Year 2	Male	29	1 years
R5	Resident Year 2	Male	29	1 years
R6	Resident Year 3	Male	33	2 years

on the study context, and requires uniquely-designed evaluation according to the task and user group. Specifically, selecting the particular type of interaction tasks (*e.g.,* drawing, pointing, and text input), as well as prioritizing speed, accuracy, or ubiquity can inform the choice of modality [34]. Age and experience level can also influence user drawing performance: older adults perform significantly slower in both touch- and mouse-based tasks, yet the lesser-familiar touchscreen interactions reduce the gap compared to younger adults [27].

Unlike prior research that focused on general tasks and target users, this paper offers empirical evidence of the impact of different delineation modalities and input devices within realistic contouring tasks with radiation oncologists. This is important given how contouring is significantly more complex than a general drawing task used in many of the studies reported above, as it involves high cognitive load [9] and errors that can lead to detrimental outcomes for patients [48, 56]. We first conducted ten interview and think-aloud sessions with physicians to contextualize contouring and examine common practices. We then followed these insights to design a cross-device prototype for contouring, engage eight clinical residents in a within-subject lab study comparing the use of Desktop & Mouse, Tablet & Stylus, Tablet & Finger, and Phone & Finger for contouring tasks, and report time spent, accuracy, and usability scores. Our results add to the large body of interactive technology literature around input and interface designs with a high-stakes and multi-faceted clinical delineation task. Exploring the task of contouring complements the existing work on input modalities, and introduces key design aspects that can inform other cross-device solutions in healthcare.

3 NEEDFINDING: EXPLORING CONTOURING PRACTICES

This section describes interview and think-aloud sessions conducted with faculty and residents to examine the existing contouring practices, types of devices commonly used, and set-up strategies.

3.1 Methods

Four radiation oncology faculty and six residents – from the Department of Radiation Medicine at the UC San Diego Health

System – participated in this study, as is shown in Table 1. While this work examined the medical context in the United States, this type of residency (and the involved clinical tasks) is common in other radiation oncology programs [23, 26]. We conducted ten remote interviews, recorded and transcribed all sessions, and analyzed the data according to video and thematic analyses. Institutional Review Boards (IRB) approved the study protocol.

Study Design
1) Faculty Sessions— Four radiation oncology faculty participated in one-hour interview sessions, consisting of two activities:

(1) In the first half of the sessions, the faculty participated in concurrent think-aloud studies [54]. The participants demonstrated a typical contouring session using their preferred software and medical case. They also expressed their thought processes out-loud: the faculty mentioned how they set up contouring sessions, what images they used, where in the screen they looked, and how they felt about their decisions. We minimally interrupted, only when the participants had not spoken for a while. We also took notes of key events and observations.

(2) In the second half, semi-structured interviews started by asking clarifying questions about our observations in the think-aloud study. These questions aimed to reveal the faculty's routine workflow about contouring practices, including the following guiding questions:
- what device(s) and software(s) do you use for contouring?
- what are the benefits and challenges when using your preferred contouring device and software?

2) Resident Sessions— Following the faculty workshop, six residents participated in one-hour sessions with two activities:

(1) The residents first filled out a brief survey on their background information (*e.g.,* age and prior medical school) and contouring practices (*e.g.,* main contouring devices).

(2) The residents then performed retrospective think-aloud [54], given that a retrospective approach can lessen the cognitive load of residents (who are less experienced than faculty) and enhance simultaneous verbalization and task performance [53]:
- In the first half, the participants contoured a case of their choice without narration. They shared their screen as the researchers recorded the silent video feeds on their local machines. The researchers also logged notable events and avoided asking questions.
- In the second half, the participants retrospectively watched their contouring sessions and explained thought processes and decisions. The researchers granted remote control of their computers to enable residents adjust video speed and playback (*e.g.,* residents could pause the video when they sought to provide long explanations).

Analysis
1) Video Analysis— Analyzing the contouring portions of the faculty and resident interviews aimed to reveal common contouring practices and patterns. Two researchers iteratively coded and discussed the videos using ChronoViz [29]: first, one researcher coded one video and extracted a collection of common actions. The other researcher then used these actions to code another video and modified the actions, as needed. The two researchers then discussed the codes and converged on a set of actions and definitions that fit both videos. This iterative process repeated five times for each pair of videos. The final set of actions guided re-coding of all videos.

2) Thematic Analysis— Inductive (*i.e.,* bottom-up) thematic analysis [13] of faculty and residents' interviews contributed to understanding broad contouring behaviours and feedback exchange mechanisms. The first author open-coded the transcribed interviews and identified the main topics. Iterative discussions among the team merged these initial codes into preliminary, and then, final themes.

3.2 Results

This section presents three main themes on contouring practices. The reported distribution percentages emerged from examining instances of the relevant actions in all sessions. F1–F4 represent the four faculty, and R1–R6 refer to the residents, as shown in Table 1.

1) Action Types: Participants delineated and navigated frequently and successively
The video analyses of faculty and resident sessions revealed three main categories of contouring behaviours (as displayed in Table 2): Set-up, Delineation, and Navigation.

Category 1: Set-up— At the beginning of every contouring session, the participants set up a workspace layout for viewing multiple images simultaneously. The most common layout (depicted in Figure 3a) contained one main window on the left with two smaller windows stacked vertically on the right. Few participants instead selected two equally sized windows adjacently. Other observed actions in the Set-up category were defining contouring structures, adjusting image properties (*e.g.,* brightness and contrast), and zooming. Some participants avoided enlarging regions of interest, such as F4 who explained: *"[zooming] would just make the image too pixelated, and I wouldn't be able to see the boundary of what I'm contouring"* (F4). In addition to the main contouring tool, R2 pulled up the patient consultation notes, and placed it in a separate monitor. As R2 further explained, these notes described the patient's prior treatments.

Category 2: Delineation — This category consisted of placement and refinement of the curved, enclosed contours. All participants delineated axial slices in two steps. *First*, they quickly defined an outer edge of their area of interest by either tracing around the region (200 instances out of 238; 84.0%) or duplicating the previous contour on the adjacent slice (38 instances; 16.0%). *Second*, before continuing to the next slice, the participants fine-tuned the contours: using the same drawing tool, they pushed the erroneous edges inward or outward. During this process, the participants frequently adjusted the thickness of the drawing tool according to the region size. The participants also used three automatic delineation mechanisms: expanding contours by a given margin, filling in empty slices using linear interpolation of nearby contours, and smoothing jagged edges. Most participants preferred smoothing their contours primarily due to aesthetics, while some believed it can simplify subsequent dose calculation for large areas:

> *"It's more appealing to see a smooth contour. It probably doesn't make a huge difference in terms of where the radiation dose is going. But I do know that in a prostate case, it can help with dose calculation."* (F3)

Category 3: Navigation — The participants navigated through medical images to either initiate the next contouring or evaluate the 3D anatomical structure. Retrieving the next target axial slice for contouring frequently occurred in all sessions, yet the participants followed different strategies: while most residents and faculty manually contoured all the slices (144 instances out of 236; 61.0%), some participants skipped one slice (75 instances; 31.8%) and used the interpolation tool to automatically compute the contours for the skipped slices. In few instances (17 instances; 7.2%) the participants skipped two or more slices before contouring the next slice. As explained by R4, the rate of change in the anatomy of nearby slices informed the strategy for acquiring the next target slice:

> *"Initially I was skipping one slice. Sometimes I'll skip a couple of slices if the anatomy isn't changing very much between slices. Towards the end where things were changing more quickly, I contoured on those last few slices manually every slice. I could just interpolate it and then fix it. But, because it's changing so quickly, the interpolation probably won't work very well. So I'll just have to fix it manually anyway."* (R4)

Reviewing slices was another type of action in the Navigation category. All participants rapidly viewed slices in succession, both in the axial and sagittal planes. Most reviewing instances occurred in the axial plane (187 instances out of 196; 95.4%), while in rare instances the participants also reviewed the sagittal slices (9 instances; 4.6%). While the participants did not navigate through the coronal image plane, some mentioned that looking at this view *"helps to see where [they] were going in all three dimensions"* (F2). Understanding the general anatomy was the main reason to consecutively view all slices. The participants also reviewed slices to *"make sure the image quality [was] reasonable"* (F4), *"compare with the guidelines"* (R6), and *"check interpolation"* (F3).

Many delineation and navigation actions occurred jointly in similar patterns. As depicted in Figure 2, most fine-tuning appeared immediately after outlining when contouring a single slice. Single-slice navigation also followed each set of delineation actions: the participants navigated to the next target slice after finalizing contours on the current slice. Reviewing the axial slices after interpolation was another common pattern which enabled participants to evaluate the accuracy of automatically generated contours and refine if needed.

2) Input Devices: Trackpad owners favored contouring using stylus over the default mouse input

Category	Action	Definition
Set-up	Setting layout	Placing images in pre-defined layouts for simultaneous viewing
	Defining structure	Assigning name, type, and color for new contouring structures
	Manipulating images	Adjusting the brightness, contrast, and opacity of images
	Zooming	Enlarging certain regions of images
Delineation	Outlining	Drawing rough contours in one stroke
	Auto-placement	Placing duplicated contours from the adjacent slices
	Fine-tuning	Altering the existing contour by pushing edges inward or outward
	Resizing	Adjusting thickness of the drawing tool
	Expanding	Adding margins to existing contours
	Interpolating	Generating contours automatically via linear interpolation of the adjacent contours
	Smoothing	Flattening the jagged edges of contours
Navigation	Viewing adjacent slice	Switching to the immediate neighboring slice (*i.e.,* viewing ± 1 slice)
	Skipping one slice	Switching to the next-but-one slice (*i.e.,* viewing ± 2 slice)
	Skipping two or more slices	Switching to at least the next-but-two slice (*i.e.,* viewing $\geq \pm 3$ slice)
	Reviewing (in axial)	Changing slices continuously in the axial plane
	Reviewing (in sagittal)	Changing slices continuously in the sagittal plane

Table 2: Common contouring actions as observed in the interview sessions, categorized into Set-up, Delineation, and Navigation.

In their daily practice, most participants contoured using the keyboard and mouse set-up while viewing images on external monitors, yet they recalled alternative input methods and devices from prior experience. Some faculty also owned a separate trackpad and stylus (*i.e.,* digital pen) that they connected to their desktop: F1, who contoured using a stylus and trackpad during the interview, favored this method especially when delineating *"lots of bigger structures with sweeping shapes"* (F1), and further elaborated that this method produced *"smoother lines"* (F1) and a *"faster and more satisfying"* (F1) contouring experience. Despite these benefits, F1 admitted that he used the mouse for simpler contouring tasks due to the *"activation barrier to switch from the mouse [used for non-contouring tasks] to the pen"* (F1). F4, who also owned a trackpad, mentioned the lack of accuracy with a mouse device: *"you don't do any precise work with this shape of your hand [shows hand gesture holding a mouse]. You do it like this [shows hand gesture holding a pen]"* (F4). While the other two faculty did not own a separate trackpad and stylus, they expressed that with more regular use, they would feel more comfortable using these devices. For instance, F2 reflected on her previous experience:

> *"I've tried the stylus. I need to use it on a more regular basis to be comfortable with it. And since I don't own one myself, I just haven't made that jump."* (F2)

Contouring directly on a specialized touch screen (*i.e.,* large display tablet compatible with the existing contouring software [35]) was another strategy. Yet, none of the participants personally owned such hardware, and had only tried devices provided by their institutions. These participants favored this input method the least and identified the fat-finger phenomenon as the main drawback of direct contouring on a touch screen: *"I used [touch screen] during my residency. I often have my hand get in the way of what I was trying to look at on the screen"* (R1). F4 echoed this challenge and raised discomfort with his hand orientation using this method: *"I don't like the touch screen, because my hand blocks my view of the images and I find it tedious to be holding my hand up on a screen anyway. It's tiring"* (F4).

3) Combining Images: Simultaneous viewing of different images assists region identification

The faculty and residents strongly preferred simultaneously accessing different types and planes of images to inform identifying regions. As explained by the participants, the radiation dose is calculated based off contours only on the CT scans. However, not all regions appear distinctly on CTs: as such, other types of medical images (*e.g.,* PET and MRI) can assist contouring. F1 – who contoured a brain case – explained his decision for selecting both MRI and CT:

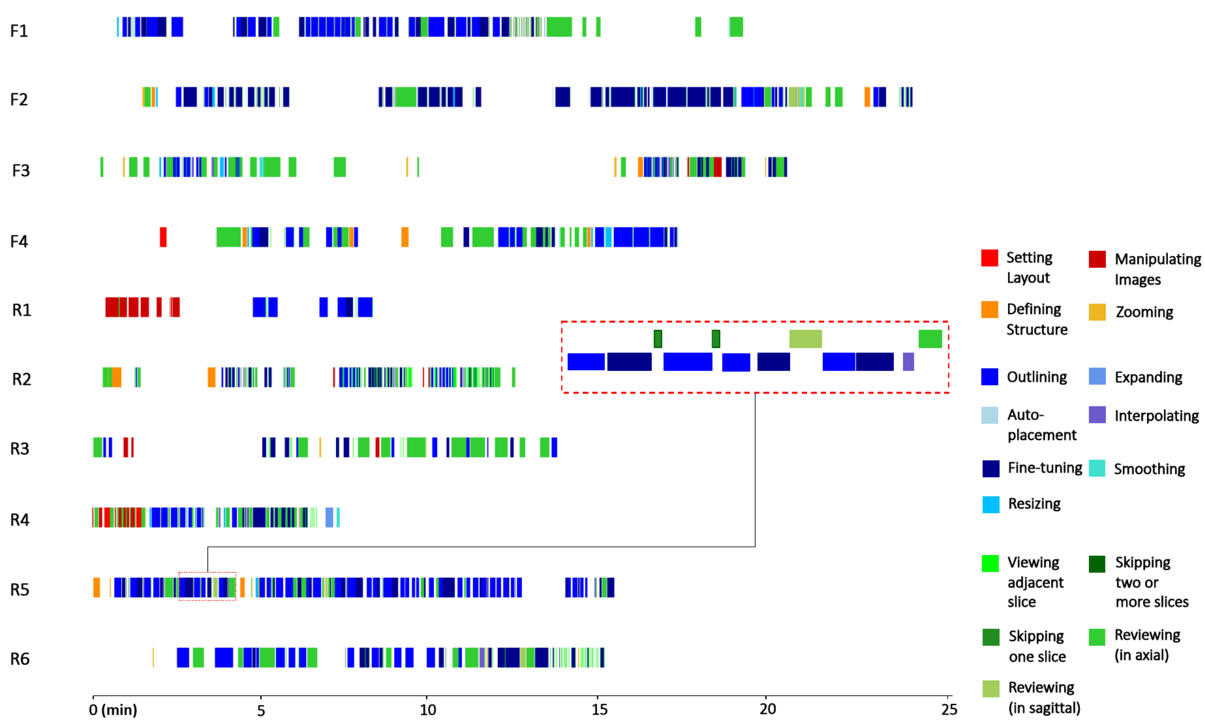


Figure 2: Recorded contouring actions during the interviews which took between 15-17 minutes (residents) and 20-25 minutes (faculty). White spaces indicate speech (*e.g.,* explaining medical concepts) and miscellaneous actions (*e.g.,* consulting external resources). An enlarged, two-minute snapshot of R5's session exemplifies common contouring patterns: R5 fine-tuned slices immediately after outlining. He also skipped slices to later use interpolation. The final reviewing aimed to assess the quality of the interpolated contours.

> *"For stuff in the brain CT is not that great. [CT image] is just a very similar gray color. And so we always use an MRI for treating things in the brain. For other parts of the body PET scans can be really useful."* (F1)

Another key factor in image selection was the anatomical plane. All participants benefited from concurrently viewing images in the sagittal plane while contouring in the axial plane. F4, for instance, expressed that he was looking at the sagittal view *"just to make sure that it was anatomically plausible"* (F4). He later elaborated:

> *"I usually start with just contouring on axial. Glancing over at the sagittal, if it doesn't look plausible, then I will look at why and see if I made a mistake."* (F4)

To facilitate concurrent viewing of medical images, the faculty and residents either placed these scans side-by-side or fused them. The contouring tools supported adjacent placement of images via a series of pre-defined layouts (two examples depicted in Figure 3): the most common layout of the participants' contouring sessions was a half rectangle on the left for the main contouring (axial plane which displays slices in the orientation of head to tail) and two quarter rectangles on the right for alternative views, such as sagittal (*i.e.,* left to right) and coronal (*i.e.,* back to front). Some participants instead chose to fuse the images. Image fusion refers to overlaying two pre-registered images with adjustable opacity (Figure 3b). The choice between displaying images adjacently or fused mostly came down to personal preference. For instance, F3 favored image fusion: *"I just blend the two images, which I personally think is more helpful than it would be to put them side by side"* (F3), while F1 preferred side-by-side viewing as *"it's just much more efficient. And you can actually contour on either one"* (F1). F4 – who used both methods – attributed his decision to the registration accuracy:

> *"If it's a PET/CT, meaning the PET and the CT were done the same time, then I want them overlaid. But if the PET is not the same as the CT that I'm drawing on, I like to have them side by side because, when they're side by side, I can contour on either one. But I can also see where the registration might not be right."* (F4)

3.3 Design Implications

As pointed out in the Results section, while mouse is the default input modality for contouring (due to entry barrier of purchasing new hardware and existing familiarity with a mouse device), few participants – who owned touch devices – benefited from *"smoother lines"* (F1) and *"faster"* (F1) contouring experience of trackpad and stylus. The Results section also unveiled a common delineation technique via successive rough outlining and precise refining, and further showed the benefit of additional screen space for simultaneous viewing of different images. These findings point to the potential of task division between touch- and mouse-enabled everyday devices, as described in this section.

Cross-device Interface—In a multi-device contouring ecosystem, radiation oncologists can overview and roughly place contours using their readily-available touch devices (*e.g.,* phone and tablet), and later, fine-tune these contours on more specialized equipment (*e.g.,* desktop computers with external monitors). Prior works support this division of tasks and input modalities: in tracing curved shapes with a single stroke, finger touch produces the lowest completion time and highest error (compared to stylus and mouse inputs), while stylus yields the highest accuracy [61, 63]. More broadly, gestural interfaces (enabled by touch devices) can offer performance benefits over conventional mouse-keyboard systems due to lower attentional demands, easier mapping of intentions, and lesser operational steps [55]. However, as noted by F2, physicians' familiarity with the existing workflows, as well as the training needed to adapt to a new technology, should be considered in designing effective computer-supported tools in healthcare.

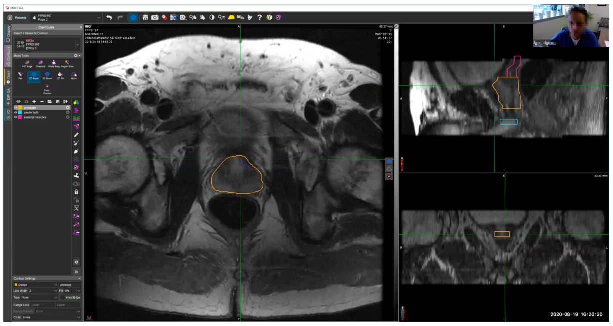
Cross-device Interaction—Contouring on a touch device with simultaneous viewing on external monitors can enhance the overall delineation experience by capitalizing on the affordances of cross-device workflows [15, 50]. Despite potential benefits of contouring on touch devices, a considerable challenge is the fat-finger phenomenon [12, 51, 57], as also mentioned by the participants. This is especially important in contouring, since physicians need to examine the subtle imaging patterns. Besides, radiation oncologists frequently set up layouts of different image types (*e.g.,* PET) and views (*e.g.,* sagittal) when having access to large screens, as noted in the Results section. An additional advantage of large (or multiple) monitors is facilitating instantaneous access to less frequently used resources [32], such as consulting other images and patient notes. This cross-device interaction (*e.g.,* contouring via touch connected to external monitors) can eliminate the fat-finger problem and facilitate viewing multiple planes and types of images.

The presented cross-device strategies can provide flexibility and enable physicians to adjust contouring systems to their needs. As explored in many socio-technical works, users not only adapt to systems, but also adapt systems to their needs [45, 47]. The proposed contouring ecosystem should anticipate various integration of navigation and delineation steps (*e.g.,* outlining, skipping), as well as input and device modalities (*e.g.,* mobile and touch, desktop and mouse).

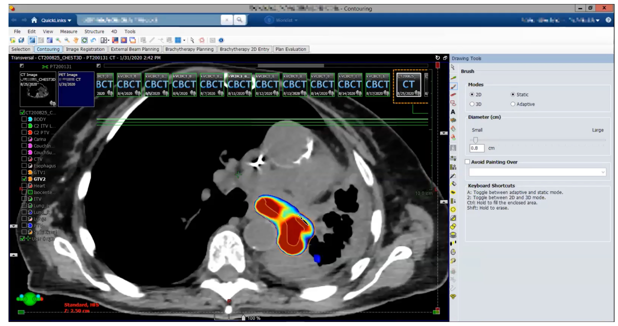
4 ICONTOUR: CROSS-DEVICE CONTOURING PROTOTYPE

To assess the feasibility of contouring based on cross-device design implication, we designed a proof of concept interface that leverages core contouring actions and common behaviours (as described in Sec. 3.2) and presents a simple, yet effective interface that is compatible with desktop, as well as touch-based devices. To ease physicians' adaptability to this prototype (given their established familiarity with existing workflows), this interface follows a button-based design which incorporates the core contouring features, as laid out in Table 2. This section describes the design and development of the prototype (Figure 4) according to the main components: the canvas area, navigational mechanisms, and control buttons.

Canvas Area: The largest section of the interface (as shown in Figure 4a) contains two overlaid `HTML Canvas` elements which facilitate viewing images and delineating contours. The `Cornerstone API` [1] facilitates loading and displaying the



(a) A faculty's contouring session using MIM software [3]. F4 used a three-image set-up with three orientations of the same MRI images. He placed contours on the axial slices (left) and assessed the anatomical structure using the other planes.



(b) A resident's contouring session (using Eclipse [2]). R4 contoured and viewed only axial slices that displayed fused CT (gray areas) and PET (bright regions) images. He used the bottom, horizontal lever to adjust the opacity of the two sets of images.

Figure 3: Two screenshots of the recorded interview sessions. The identities of physicians and patients are anonymized.

256×256 cross–sectional images which are previously extracted from anonymized DICOM (Digital Imaging and Communications in Medicine) files of patients' CT scans and MRIs. The second overlapping Canvas places user contours on top of the images using the `Paper.js` framework [4] which aids creating and manipulating vector graphics. Physicians can zoom the images by two-finger pinching and expanding (on touch devices) or using the scroll-wheel on mouse devices, which are common interaction techniques in image viewing and navigation/map applications.

Informed by the frequent and successive contouring patterns of outlining and refining (Figure 2), the delineation tool eases switching between the two modes to expedite the process (Figure 5). When users first delineate a slice, the coordinates of the brush contribute to an outline of the contour (i.e., outlining mode). Subsequent user strokes, if intersected with existing contours, push the structure inwards or outwards by subtracting or adding the new stroke coordinates from the existing contour. Besides, to address the potential fat-finger problem [51], the indicator brush displays a circle that dynamically changes size to encompass slightly larger than the surface area of the touch input (*e.g.,* finger). Given that the outer edge of the entire stroke (formed by continuous brush circles) shapes the overall contour, this technique can enable physicians to view and follow along their contour from outside their finger, mitigating the fat-finger problem.

Navigational Mechanisms: The contouring prototype contains three methods for navigating images, displayed in Figure 4b. *First*, four vertically-stacked, large buttons facilitate switching the displayed image to the immediate neighboring slices (*i.e., prev 1* and *next 1* when contouring all slices), and next-but-one slices (*i.e., prev 2* and *next 2* when skipping slices). The large size and close proximity of these buttons to the main canvas can expedite back and forth delineation and navigation, as laid out in Sec. 3.2. *Second*, the vertical slider near the edge of the interface enables fast viewing of successive slices which can serve two purposes:

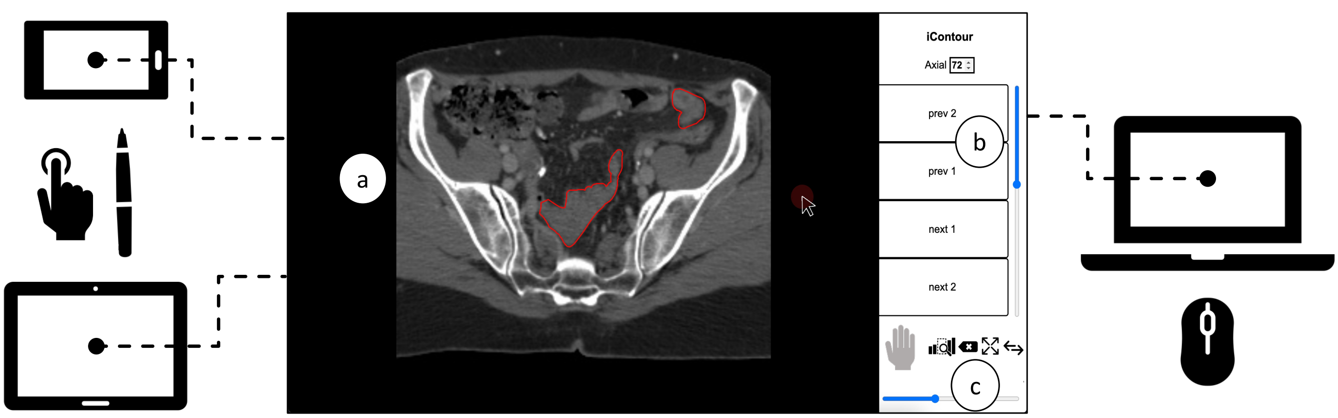


Figure 4: The cross–device contouring prototype that enables examining and delineating medical images on desktop, phone, and tablet via different input modalities like mouse, stylus, and finger. (a) The canvas region contains the medical image and facilitates outlining and refining of contours. (b) The four *prev* and *next* buttons, as well as the vertical slider and the input text box (which displays the slice number) provide both coarse- and fine-grained methods for navigation and target slice acquisition. (c) These buttons offer image and interface manipulation functionalities: from left to right, panning, interpolation, clearing all contours, screen maximizing, toggling the panel position, and a slider (in the bottom) that adjusts the brush size for delineation.

before contouring sessions, physicians can overview the slices to understand the case and the existing regions of interest. Later, the slider can help users review and validate their contour according to the broader human anatomy. *Third*, the displayed slice number is editable, meaning users can type in a number and quickly navigate to the corresponding slice. While this technique might not directly benefit contouring sessions, it can serve as a point of referencing and de-referencing for post-hoc case reviews.

Control Buttons: Five buttons and one slider appear in the bottom section of the panel (Figure 4c). The leftmost button is the drawing/panning toggle which allows physicians to switch between delineating and panning. Due to potential high frequency of use during contouring sessions, the interface design places this button — which is slightly larger than other buttons — right next to the canvas. This can improve target acquisition by decreasing distance and increasing size, according to Fitts' Law [41].

The interpolation button — which appears to the right of the toggle button — interpolates the existing contours to the skipped slices. As described in Sec. 3.2, interpolation is a common contouring technique for large, slow-changing regions which results in considerably faster operation than contouring every slice from scratch. The interpolation algorithm incorporated in this tool first pulls contour coordinates from existing slices (stored on the server as Scalable Vector Graphics, or SVGs) and converts them into binary images. The `OpenCV` library [5] then expands points to fill large gaps in the coordinates files. A morphology-based contour interpolation pipeline then interpolates these coordinate files over the skipped slices. Specifically, we used Albu *et al.*'s implementations of inter-slice interpolation [6] due to its effectiveness on creating gradual change of shape without generating over-smoothed regions. The effectiveness of this algorithm has also been validated in prior contouring systems [18]. Lastly, the interpolated contours are converted back into points and the final SVGs are stored on the server. Two additional mechanisms are introduced to support specific delineation mechanisms. The middle button in the control panel clears all contours on the current slice. While the brush tool can fully erase (i.e., in refining mode when the stroke covers the entire existing structure), this button can be a faster method to re-contour an entire slice. The horizontal slider (below the five buttons) controls the radius of the brush from 5 to 50 pixels, given that *resizing* is a common contouring action, described in Table 2.

Finally, the two rightmost buttons facilitate customizations for the size and placement of UI elements. To fully utilize the screen space of small touch devices, a full-screen button fixates the orientation of the interface to landscape mode and hides the URL input space of the browser. The rightmost button in the section moves the entire panel symmetrically between right and left sides of the interface. The symmetric shifting enables all components to stay consistently proximate relative to the main canvas region: for instance, the four navigational buttons stay the closest to the image on either side of the interface (compared to the vertical slider) to expedite back and forth delineation and navigation.

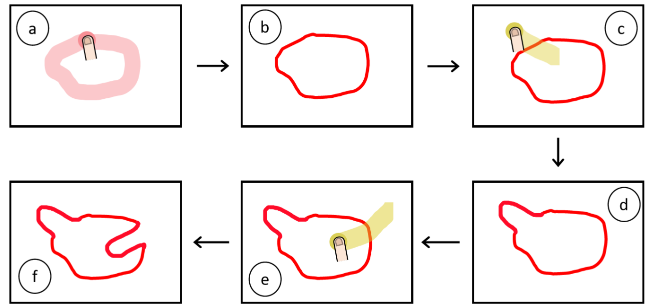


Figure 5: The design technique for fast and accurate outlining and refining on touch devices. (a) The user's first stroke on an image constitutes *outlining* and appears with low-opacity red (for the traced path) and a high-opacity red circle (for the position of the input modality). (b) Once the user disengages the input modality (*e.g.,* lifts finger), the resulting contour appears with a red opaque stroke. (c) The user can *refine* the contour from inside to outside with a yellow low-opacity stroke describing the traced path. (d) The contour has a part protruding out which marks the outline of the refined path. (e) The user can also *refine* the contour from outside to inside. (f) The contour, now, has a part hollowed out.

5 LAB STUDY

We evaluated the iContour prototype in a lab study at UC San Diego Health, in order to understand the feasibility of contouring across desktop and everyday mobile devices. In a within-subject study, participants performed contouring tasks using four randomly ordered conditions: Desktop & Mouse, Tablet & Stylus, Tablet & Finger, and Phone & Finger. To facilitate consistent comparison between conditions, the algorithm-dependent interpolation tool was disabled from the interface. Given that the focus of the study was to examine the feasibility of different device and input modalities, comparing time spent contouring can be challenging if some (and not all) participants choose to interpolate contours. The Institutional Review Board (IRB) approved the study protocols which included written consent forms to record sessions.

5.1 Participants

Eight radiation oncology residents (out of 10 overall at the residency program of UC San Diego Health) participated in the lab

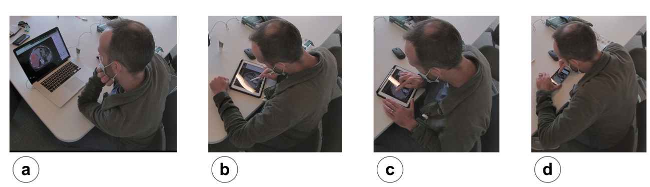


Figure 6: Four snapshots of the within-subject lab study with radiation oncology residents. Each participant performed the same contouring task on (a) Desktop & Mouse, (b) Tablet & Stylus, (c) Tablet & Finger, and (d) Phone & Finger.

study. The participant pool comprised four females and four males, aged between 25 and 40 years old. The participants were in the second to fifth year of their residency programs with varying, yet sufficient levels of contouring experience. The relatively small number of participants reflects the highly specialized domain of these physicians: a 2017 report shows that the American healthcare has an average of 1.64 oncologists per 100K people [11].

5.2 Procedure

During the study, the residents took part in the following four steps:

[5 mins] On-boarding: The participants first learned about the goal and procedure of the study. Then, the lead researcher explained major components of the interface and demonstrated these features on the desktop and touch devices. The consistent functionality and aesthetics of the interface enabled fast explanation of the components.

[10 mins] Learning: The lead researcher then laid out all devices (laptop, phone, tablet, and stylus) and encouraged the participants to try out all four conditions. The participants also learned about the broad topic of the task (i.e., contouring the liver) which prompted exploration of the case and locating the liver in the images. An easily identifiable and well-defined structure, such as the liver, accommodated varying levels of resident experience. The relatively long learning phase aimed to lessen the impact of carry-over effects in a within-subject study [31].

[32 mins] Contouring: The main contouring task consisted of four, eight-minute long segments, one for every randomly-assigned condition. The lead researcher explained the task description in the beginning of each segment: *"please contour the liver in slices 120-130 in the next, at most, 6 minutes."* The wording aimed to create a practical balance between accuracy (*i.e., "contouring"* by definition) and speed (*i.e., "at most 6 minutes"*). After each task, the participants completed the 10-questions System Usability Scale (SUS) survey [14], which took additional 2-minutes per segment. These two steps repeated four times in total. The contouring sessions were recorded in two ways: first, a wall-mounted camera captured participants' external interactions with devices, such as holding the phone and tablet. Second, the content of the screen in all four conditions was recorded via screen-sharing on a video conferencing tool.

[15 mins] Interviewing: At the end of the session, the lead researcher conducted an exit interview to gauge each participant's impressions on the following questions: (1) *how did you feel about the overall experience*, (2) *how do you rank the four conditions you tried today*, and (3) *how would you adopt cross–device contouring in your medical practice.*

5.3 Data Collection and Analysis

This study captures the following metrics and analyzes statistical difference according to Repeated Measure Analysis of Variance (RM-ANOVA) ($\alpha = .05$) and pairwise Fisher's Least Significant Difference (LSD):

Accuracy: The Dice Similarity Coefficient (DSC) evaluated the correctness of the produced contours. DSC — a popular measure in radiation oncology training [25] — gauges the similarity between two structures with a score from 0 (no overlap) to 1 (perfect overlap). One of the participating experts (author) provided the gold standard contour on the same liver case — which possesses a well-defined structure — as the point of comparison in DSC calculations.

Time Spent Contouring: The recorded video screens provided accurate timing on contouring sessions, including how long participants spent contouring every slice.

SUS score: The System Usability Scale questionnaire, which the lead researcher administered at the end of every condition, contained 10 usability questions. The consolidation of these questions leads to a score between 0 and 100, qualitatively defined as *worst imaginable* and *best imaginable*, respectively [10].

The Results section further reports notable contouring actions (captured via wall-mounted camera and screen recording), as well as key quotes expressed during the exit interviews.

6 RESULTS

The lab study marked the first time that any of the participating residents performed contouring on a tablet or phone, as self-reported in the interviews. Many participants were surprised to be able to contour on an everyday tablet, and especially, a phone device. This section first presents the overall perception of the design techniques and general iContour UI (Sec. 6.1). Sec. 6.2– 6.4 then describe findings in terms of accuracy, time spent contouring, and system usability [59].

6.1 iContour Satisfied Contouring Needs and Prompted Unique Patterns

Overall, the participants carried out the contouring tasks without significant challenges on all four devices and input modalities, despite the fact that none had prior mobile–based contouring experience. Some were even surprised to be able to contour on small touch devices such as a phone: *"contouring on the phone was easier than I thought it was going to be. I thought it would be hard because it was so small, but definitely easier than I thought"* (P4).

Most participants expressed that the design of the prototype satisfied their contouring needs and requirements. They managed to sufficiently learn the UI components (within the first few minutes of introduction), and complete the contouring tasks accordingly. Some participants commented that the choice and placement of menu items facilitated contouring and eased repeated use of the provided affordances:

> *"I felt I was getting into a routine to start with a bigger brush and then refine with a smaller brush. On either one of the four conditions, it wasn't really a problem going to the menu and coming back."* (P3)

A few of the participants, however, struggled to get used to the contouring functionalities: *"it was a little awkward, because it is very different than the system I am used to contouring. It would just require more practice"* (P5).

The screen recordings revealed common contouring patterns on all conditions: the participants first zoomed and panned the image to capture the entire region of interest (*i.e.,* liver in this case), and then consecutively went through slices and placed contours. The frequent use of brush size slider enabled repeated coarse outlining and fine-grained refining. Some participants chose to continuously adjust the brush size when contouring every slice, while others performed multiple rounds and kept the brush size consistent in every round: these residents still panned and zoomed right before performing a round of delineations. This latter behaviour might have developed to address the additional effort needed to adjust brush size using the dedicated slider.

The wall-mounted cameras captured the participants' interactions with the devices. All participants, who were mostly right-handed, placed all devices directly on a supporting desk to stabilize their contouring sessions (see Figure 6). These residents performed delineation and navigation actions of touch devices using the index or middle fingers of their dominant hands. The only left-handed participant of the study decided to place the navigation and control panel on the right side: this resident delineated slices with their left hand, but used the navigational buttons and slider via their right hand, and further highlighted satisfaction with dividing tasks between their dominant and non-dominant hands: *"it was nice to have my non-dominant hand scroll through images, while my left hand contoured"* (P7).

6.2 Tablet & Stylus Produced the Highest Accuracy

When measuring accuracy — most critical criterion for safe contouring — comparing participant contours to a baseline gold standard showed similar results across all devices, though the Desktop & Mouse yielded a slightly lower DSC (0.916), and the most accurate was the Tablet & Stylus (0.923). As displayed in Figure 7, the other finger-based conditions yielded 0.922 (Tablet & Finger) and 0.917 (Phone & Finger) DSC scores. The variability between conditions, while statistically significant ($F(3, 348) = 18.84$, $p < .0001$), are likely not *clinically* significant: as deemed by the participating experts, the small difference of accuracy across conditions might not lead to significant clinical outcomes, such as patient survival.

6.3 Phone Yielded Fastest Contouring per Slice

As shown in Figure 8, the participants spent the least amount of time contouring on touch devices, specifically in the Phone & Finger condition with 18.6 seconds, followed by the two tablet-based conditions with 19.2 (Tablet & Finger) and 21.5 seconds (Tablet & Stylus). Surprisingly, Desktop & Mouse — which is the default device and input modality in clinical contouring — resulted in the longest contouring time of 26.8 seconds per slice. RM-ANOVA deemed differences across groups significant ($F(3, 348) = 6.82$, $p < .001$).

6.4 Participants Highly Favored Tablet & Stylus

The SUS questions revealed strong preferences for contouring on Tablet & Stylus with a score of 73.13, followed by Tablet & Finger which achieved a score of 68.44. While still within range of "acceptability" [10], Desktop & Mouse and Phone & Finger produced much lower usability scores of 56.88 and 54.06, respectively. RM-ANOVA revealed statistically significant

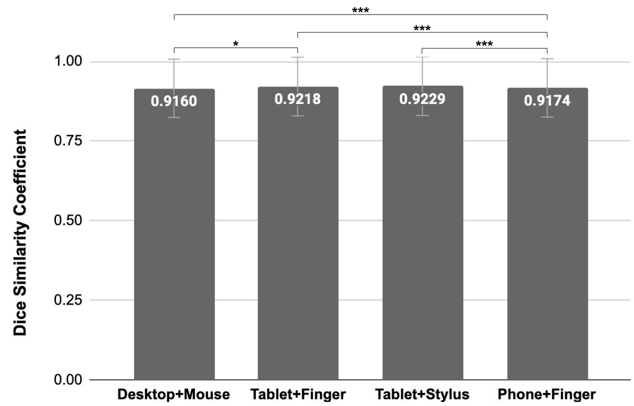


Figure 7: The accuracy of the produced contours measured in comparison to gold-standard contours via DSC. Tablet & Stylus produced slightly higher accuracy compared to the other conditions ($p < .0001$), though the differences are clinically deemed insignificant.

differences among the four conditions ($F(3, 28) = 5.41$, $p < .01$), as shown in Figure 9.

When asked to rank the four conditions, all participants placed Tablet & Stylus in their top two favorites, pointing out that they felt *"most precise"* (P6) and *"most confident"* (P3), and that it was *"much easier to control"* (P2) and *"really amenable to very small details"* (P8). P1 — who had never used a tablet and stylus for contouring — highlighted the need for contouring on everyday personal devices:

> *"I found the most fluidity with the iPad and stylus. I actually really liked it. There are really expensive, commercial, stylus-based workstations for contouring. As residents, we mostly use desktop and mouse."* (P1)

Most participants highlighted that the Phone & Finger condition was their least preferred, mainly due to small screen

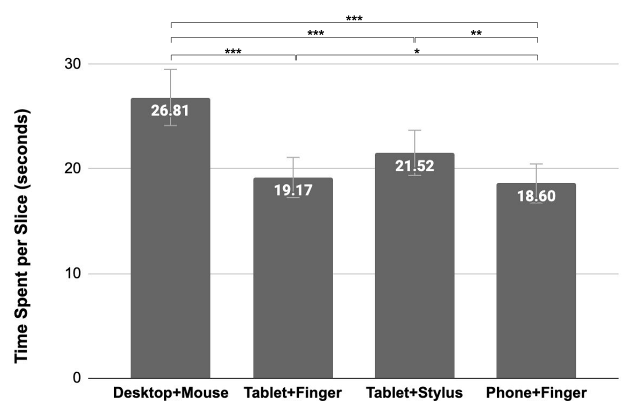


Figure 8: Average time spent contouring per slice for each condition. Desktop & Mouse led to significantly longer times than touch devices. Phone & Finger, despite the smallest size, facilitated the fastest contouring. Overall distributions are statistically significant ($p < .001$), and post-hoc pairwise tests show statistical differences across most pairs.

space. Compared to a larger sized tablet, which received a considerably higher usability score, the participants suffered from lack of visual context from the images:

> *"I liked having a bigger screen. On the phone, when I got the liver to be big enough where I felt comfortable contouring with finger, I didn't necessarily have the entire picture anymore. So I had to repeatedly make it smaller, move it around, and then expand it."* (P6)

Some participants, however, highlighted the potential for granular and ad-hoc feedback on mobile devices:

> *"If you wanted to talk to somebody or show them something, you could easily pull out your phone when you run into them. So they could quickly adjust your contours. [With the current tools] you can't really review contours unless you are both sitting down at a computer, and there is a time set for it."* (P4)

Surprisingly, the residents' default contouring method (i.e., Desktop & Mouse) yielded relatively low usability scores. Some participants expressed struggling with this condition since some features differed from their own settings. One particular feature was mouse sensitivity: while some participants mentioned that the mouse used in the lab study was *"too sensitive"* (P3), others said that it was *"way less sensitive than [their] mouse"* (P8). These residents further elaborated that the former case led to *"a lot of jagged edges which [they] then needed to go back and edit"* (P3), and the latter caused *"having to move [their] hand a lot more"* (P8).

7 DISCUSSION

Overall, all participants favored cross-device contouring, and provided specific scenarios for how the additional ubiquity of everyday mobile devices can improve their current contouring practices, such as post-hoc reviewing of cases. While the results pointed to the Tablet & Stylus as the most preferred modality, the

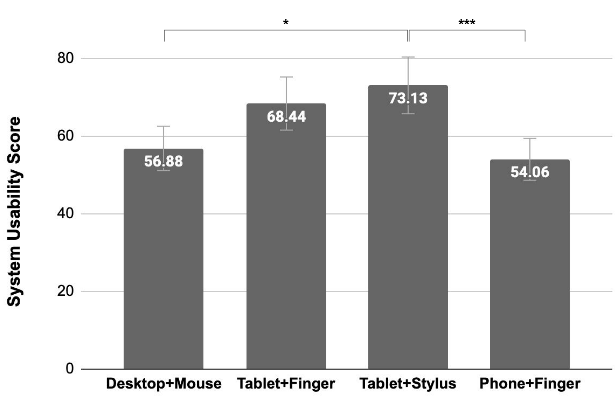


Figure 9: The average System Usability Score for every condition. Participants collectively rated Tablet & Stylus as the most usable, followed by Tablet & Finger. Phone & Finger yielded the least score. Rm-ANOVA indicated overall significant differences among conditions ($p < .01$), and post-hoc pairwise tests showed statistical mean differences between the highest and the two lowest conditions.

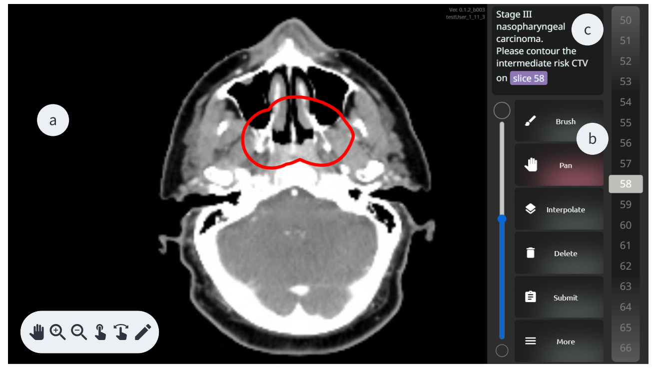


Figure 10: Re-designed UI of the cross–device contouring prototype that (a) enables delineating a stack of medical images using interactions like pan, zoom, tap, and swipe; (b) provides essential features to assist contouring, such as toggling between drawing and panning, interpolating and erasing contours, and (c) incorporates a panel in-situ of the main canvas that displays patient information.

high accuracy and low time spent of small touch devices for contouring surprised many participants. This section first reflects on the impression and contouring behaviors of the participants at the lab study to introduce a new UI of the prototype (Sec. 7.1). Then, the findings of the lab study inform exploring the potential of a contouring ecosystem in Sec. 7.2. Sec. 7.3 discusses the design techniques for cross–device contouring and presents three design guidelines for enhancing accuracy, time spent, and ubiquity in other healthcare domains. Lastly, Sec. 7.4 discusses limitations and future directions of this work.

7.1 iContour 2.0: Re-designed User Interface following the Lab Study

Reflecting on the participants' perception of the design techniques of the cross–device contouring prototype, we iterated on the aesthetics and placement of elements, as shown in Figure 10. The following describes the core changes:

- While the original four buttons and the adjacent slider provided a mix of granular and coarse navigation, these elements occupied a large space, especially on small phones. To allocate more space to other features, we instead added a scroll indicator which enables granular selection of nearby slices and fast scrolling to explore the case.
- To address some participants' struggles when getting used to button functionalities, the new UI displays titles as well as the icons which can expedite locating features.
- Given the frequent and successive delineation and brush size adjustment, the new version eases this transition by introducing a vertical slider right next to the main canvas.
- The new interface contains a text box on the top right corner that displays information about the current contouring task. Future work can use this box to provide other clinical and educational information, such as patient history and contouring hints.

7.2 Towards Developing a Contouring Ecosystem

While all four types of interactions tested in our lab study can reasonably support contouring tasks, some combinations of device and input modalities surpassed expectations and showed high feasibility. Tablet & Stylus, despite yielding only the third-best contouring time, still considerably outperformed the default Desktop & Mouse, and received the highest usability score. The residents performed the fastest using the two finger-based conditions, yet rated Phone & Finger as the least usable due to small screen space and lack of visual context. In addition, the enhanced ubiquity and convenience of carrying phone devices can create learning opportunities, in which residents can seek targeted and ad-hoc feedback. Surprisingly, not only Desktop & Mouse yielded the longest contouring time, it received significantly low usability scores. However, the accuracy of contours (the most critical metric in contouring) was similar across all device and input modalities. These results align with prior works that investigated drawing tasks: compared to touch and stylus, tracing curved enclosed shapes via the mouse resulted in the longest time and with error rates [61].

This study points to the potential of developing a contouring ecosystem, in which different configurations of device and input modalities can support contouring tasks and promote collaboration among physicians. In this environment, main contouring sessions can occur on tablets with styli, while phone devices can carry out minor adjustments, especially during impromptu meetings. For additional screen space, tablet devices can also connect to external screens, aligned with the second design implication (*i.e.,* cross-device interaction) presented in Sec. 3.3. While such ecosystem can incorporate the default desktop devices, this study did not find concrete benefits for using desktop and mouse in supporting the main contouring tasks. At least, the current physicians — who for years trained and practiced contouring on these devices — can benefit from this configuration. The combination of available device and input modalities can utilize their unique affordances to enhance the overall accuracy, time spent, and ubiquity of contouring.

7.3 Three Design Principles for Cross-Device Contouring and Healthcare Delineation

Reflecting on the participatory design and the insights from the lab study, this section presents three principles to inform the design of cross–device solutions in other healthcare tasks.

Convenience — Computer-supported healthcare systems should not only operate on different types of devices (with varying sizes and operating systems), but they should also aim to maximize ubiquity and ease-of-access, especially given the lack of time availability and technology proficiency of the specialized physicians. For instance, the browser-based implementation of iContour — instead of a native application that would have required additional setup and OS-compatibility — can build critical mass through diminishing the barrier to access [46], and further encourage ad-hoc and collaborative contouring opportunities. This is exemplified by the comment of P4 in Sec. 6.4, who perceived iContour to expedite contouring collaborations, in which residents can utilize impromptu interactions by *"easily pull[ing] out [their] phone"* (P4) to seek feedback, and adjust contours on the spot.

Consistency — Physicians should be able to switch devices without having to re-learn key functionalities of cross–device healthcare solutions, albeit sometimes resulting in nuanced trade-offs with unique and non-uniform device affordances. This principle — broadly categorized as *perceptual consistency* [24] and related to *consistency and standards* as a usability heuristic [52]— informed a primarily button-based design of features, contrary to a gesture-based approach which solely applies to touch devices. The benefits of this general principle were showcased in the short on-boarding step in the lab study, where the researcher needed to describe the general features in only one device. Ultimately, such user-friendly design approaches can lead to higher perceived ease of use and technology acceptance [22]. The one exception to this principle was the implemented zooming functionality that differed in touch- vs mouse-based interaction which is because of the well-established familiarity with these interaction techniques in many commercial systems.

Balance — The size and placement of UI elements should reflect the use patterns of the features, especially when transforming large healthcare tools (with many features that are not all equally utilized) into workflows that are compatible with small touch-based devices. As displayed in Figure 1, many manipulation mechanisms exist in specialized contouring softwares. However, creating a cross–device contouring interface required balancing features, and prioritizing ones that are more critical to contouring, especially in devices with smaller form factors. So, the implementation of the prototype focused on three main components (Table 2): image manipulation, delineation, and navigation.

While these principles guided the design and development of a contouring tool for radiation oncologists, they can also improve treatment and education in other healthcare fields. In particular, other image-based medical tasks — such as measuring tumor regions in pathology [30] and describing cavities in dentistry [39] — can benefit from more frequent analysis and annotation of images, afforded by a cross–device system that provides convenient access, consistent UI, and balanced features on everyday devices. Moreover, given the ubiquity of apprenticeship models in residency programs [44] and the limited interaction between residents and the attending faculty (due to scheduling challenges), cross–device solutions can provide further ad-hoc opportunities for feedback exchange and learning.

7.4 Limitations and Future Work

Our work engaged participants in short realistic contouring tasks and provided a familiar region for residents (*i.e.,* liver) to limit any confounding factors. We chose these examples based on the experts' recommendation, and while we believe that the results are representative of how they would interact with more complex cases, follow-up lab studies could examine longer and more complex tasks and further evaluate contouring on touch devices. Future works can also explore correlations between contouring and specific characteristics of input and device modalities (*e.g.,* screen size). In addition, while this paper showcased the feasibility of touch devices for contouring, the impact of these devices on other tasks that radiation oncologists engage with (*e.g.,* image fusion) remains to be explored. To further assess the real-world potential of cross-device contouring, follow-up works can compare each modality directly with the existing contouring software on desktop, including clinical solutions (*e.g.,* Eclipse) and experimental research tools [60].

Robust validation of cross-device contouring and a broader contouring ecosystem can benefit from further field studies with less restrictive and more diverse task designs, especially when transitioning from home to work environments. Engaging physicians in long-term deployment studies can also enable combining affordances of device configurations and reveal unique contouring patterns: for instance, future work can reveal the potential for quick outlining and interpolating (which was excluded from the lab study) on phone devices, and later refining on tablets and styli. Synchronous interactions (*e.g.,* connecting phone to monitors) can also uncover valuable insights about cross-device contouring. Field studies can also enhance familiarity with all configurations and enable personalizing device settings, such as mouse sensitivity as was mentioned by some participants.

8 CONCLUSION

This paper presents a cross-device contouring system that facilitates delineation and navigation not only on immobile desktop computers but also on everyday touch devices, such as phones and tablets. The interface and features of the developed *iContour* prototype were shaped through 10 interview and think-aloud sessions with radiation oncology faculty and residents. This work further conducted a lab study (N = 8 residents) and compared four common device and input modality configurations. Compared to the touch devices, the default Desktop & Mouse significantly under-performed in terms of time spent contouring and system usability, while it achieved similar accuracy as small phone devices. Tablet & Stylus, however, yielded the highest usability score and significantly lower contouring times.

Reflecting on the observed user patterns during the lab study, we re-designed iContour to better facilitate contouring needs of physicians. The findings further suggested that cross-device solutions have the potential to enable faster and more ubiquitous contouring without sacrificing accuracy. The greater ease-of-access and flexibility of contouring within this ecosystem can also encourage frequent re-generation of radiation plans that fit patients' changing anatomy, and promote educational collaborations among physicians. Lastly, the three design principles (*i.e.,* consistency, convenience, and balance) that we introduced can be useful for practitioners aiming at transforming static healthcare tasks into cross–device workflows that leverage enhanced mobility to improve critical treatment procedures.

REFERENCES

[1] 2021. Cornerstone Library for Javascript. https://docs.cornerstonejs.org/api.html

[2] 2021. Eclipse planning software (Varian Medical Systems, Palo Alto, CA, USA). https://www.varian.com/products/radiotherapy/treatment-planning/eclipse

[3] 2021. MIM MAESTRO (MIM Software Inc., Cleveland, OH). https://www.mimsoftware.com/radiationoncology/maestro

[4] 2021. Paper.js Open-source Vector Graphics Framework). http://paperjs.org/reference/global/

[5] 2022. openCV (Open Source Computer Vision). https://docs.opencv.org/3.4

[6] Alexandra Branzan Albu, Trevor Beugeling, and Denis Laurendeau. 2008. A morphology-based approach for interslice interpolation of anatomical slices from volumetric images. *IEEE Transactions on Biomedical Engineering* 55, 8 (2008), 2022–2038.

[7] VA Andrianarison, M Laouiti, O Fargier-Bochaton, G Dipasquale, X Wang, NP Nguyen, R Miralbell, and V Vinh-Hung. 2018. Contouring workload in adjuvant breast cancer radiotherapy. *Cancer/Radiotherapie* 22, 8 (2018), 747–753.

[8] A Aselmaa, RHM Goossens, A Laprie, S Ken, T Fechter, A Ramkumar, and A Freudenthal. 2013. Workflow analysis report. *7th Framework Programme of the European Union, FP7-PEOPLE-2011-ITN, Summer Marie Curie Research Training Network (PITN-GA-2011-290148)* (2013).

[9] Anet Aselmaa, Richard HM Goossens, Ben Rowland, Anne Laprie, Yu Song, and Adinda Freudenthal. 2014. Medical factors of brain tumor delineation in radiotherapy for software design. In *5th International conference on applied human factors and ergonomics (AHFE)*. 4865–4875.

[10] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[11] James E Bates, Akash D Parekh, Mudit Chowdhary, and Robert J Amdur. 2020. Geographic distribution of radiation oncologists in the United States. *Practical radiation oncology* 10, 6 (2020), e436–e443.

[12] Patrick Baudisch and Gerry Chu. 2009. Back-of-device interaction allows creating very small touch devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1923–1932.

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[14] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996).

[15] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. 2019. *Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices*. Association for Computing Machinery, New York, NY, USA, 1–28. https://doi.org/10.1145/3290605.3300792

[16] Neil G Burnet, Simon J Thomas, Kate E Burton, and Sarah J Jefferies. 2004. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging* 4, 2 (2004), 153.

[17] Chen Chen, Matin Yarmand, Varun Singh, Michael V. Sherer, James D. Murphy, Yang Zhang, and Nadir Weibel. 2022. Exploring Needs and Design Opportunities for Virtual Reality-based Contour Delineations of Medical Structures. In *Companion of the 2022 ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Sophia Antipolis, France) *(EICS '22 Companion)*. Association for Computing Machinery, New York, NY, USA, 19–25. https://doi.org/10.1145/3531706.3536456

[18] Chen Chen, Matin Yarmand, Varun Singh, Michael V. Sherer, James D. Murphy, Yang Zhang, and Nadir Weibel. 2022. VRContour: Bringing Contour Delineations of Medical Structures Into Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR)* (Singapore) *(ISMAR '22)*. https://doi.org/10.1109/ISMAR55827.2022.00020

[19] Chen Chen, Matin Yarmand, Zhuoqun Xu, Varun Singh, Yang Zhang, and Nadir Weibel. 2022. Investigating Input Modality and Task Geometry on Precision-first 3D Drawing in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 384–393. https://doi.org/10.1109/ISMAR55827.2022.00054

[20] Andy Cockburn, David Ahlström, and Carl Gutwin. 2012. Understanding performance in touch selections: Tap, drag and radial pointing drag with finger, stylus and mouse. *International Journal of Human-Computer Studies* 70, 3 (2012), 218–233.

[21] Indra J Das, Vadim Moskvin, and Peter A Johnstone. 2009. Analysis of treatment planning time among systems and planners for intensity-modulated radiation therapy. *Journal of the American College of Radiology* 6, 7 (2009), 514–517.

[22] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.

[23] Brock Debenham, Robyn Banerjee, Alysa Fairchild, George Dundas, Theresa Trotter, and Don Yee. 2012. 2009 Canadian radiation oncology resident survey. *International Journal of Radiation Oncology* Biology* Physics* 82, 4 (2012), 1326–1331.

[24] Charles Denis and Laurent Karsenty. 2004. Inter-usability of multi-device systems: A conceptual framework. *Multiple user interfaces: Cross-platform applications and context-aware interfaces* (2004), 373–384.

[25] Frances Duane, Marianne C Aznar, Freddie Bartlett, David J Cutter, Sarah C Darby, Reshma Jagsi, Ebbe L Lorenzen, Orla McArdle, Paul McGale, Saul Myerson, et al. 2017. A cardiac contouring atlas for radiotherapy. *Radiotherapy and Oncology* 122, 3 (2017), 416–422.

[26] Jean-Christophe Faivre, Jean-Emmanuel Bibault, Thomas Leroy, Mikaël Agopiantz, Julia Salleron, Maxime Wack, Guillaume Janoray, Henri Roché, Stéphane Culine, and Sofia Rivera. 2018. Evaluation of the theoretical teaching of postgraduate radiation oncology medical residents in France: a cross-sectional study. *Journal of Cancer Education* 33, 2 (2018), 383–390.

[27] Leah Findlater, Jon E Froehlich, Kays Fattal, Jacob O Wobbrock, and Tanya Dastyar. 2013. Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 343–346.

[28] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. 2007. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 647–656.

[29] Adam Fouse, Nadir Weibel, Edwin Hutchins, and James D Hollan. 2011. ChronoViz: a system for supporting navigation of time-coded data. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 299–304.

[30] Peter N Furness. 1997. The use of digital images in pathology. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 183, 3 (1997), 253–263.

[31] Anthony G Greenwald. 1976. Within-subjects designs: To use or not to use? *Psychological Bulletin* 83, 2 (1976), 314.

[32] Jonathan Grudin. 2001. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 458–465.

[33] Eli R Hooten and Julie A Adams. 2011. Comparing input error for mouse and touch input. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2853–2858.

[34] Ibrar Hussain, A Khan, Imran Ali Khan, Syed Sajid Hussain, and Attiqa Rehman. 2017. A survey on usage of touchscreen versus mouse for interaction. *Sci Int (Lahore)* 29 (2017), 83–88.

[35] Multi-Institutional Target Delineation in Oncology Group+ 1-210-4501127+ 1-210-4505085 fullercd@ uthscsa. edu. 2011. Human–computer interaction in radiotherapy target volume delineation: a prospective, multi-institutional comparison of user input devices. *Journal of digital imaging* 24 (2011), 794–803.

[36] Yosuke Kano, Hitoshi Ikushima, Motoharu Sasaki, and Akihiro Haga. 2021. Automatic contour segmentation of cervical cancer using artificial intelligence. *Journal of Radiation Research* 62, 5 (08 2021), 934–944. https://doi.org/10.1093/jrr/rrab070

[37] James Lamb, Minsong Cao, Amar Kishan, Nzhde Agazaryan, David H Thomas, Narek Shaverdian, Yingli Yang, Suzette Ray, Daniel A Low, Ann Raldow, et al. 2017. Online adaptive radiation therapy: implementation of a new process of care. *Cureus* 9, 8 (2017).

[38] Stephanie Lim-Reinders, Brian M Keller, Shahad Al-Ward, Arjun Sahgal, and Anthony Kim. 2017. Online adaptive radiation therapy. *International Journal of Radiation Oncology* Biology* Physics* 99, 4 (2017), 994–1003.

[39] Pasquale Loiacono and Luca Pascoletti. 2012. *Photography in dentistry: theory and techniques in modern documentation*. Quintessenza Edizioni Milan.

[40] Kalle Lyytinen and Youngjin Yoo. 2002. Ubiquitous computing. *Commun. ACM* 45, 12 (2002), 63–96.

[41] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[42] Shawna Meyer, Oryx Cohen, and Erik Nilsen. 1994. Device comparisons for goal-directed drawing tasks. In *Conference companion on Human factors in computing systems*. 251–252.

[43] Marcel Nachbar, Monica lo Russo, Cihan Gani, Simon Boeke, Daniel Wegener, Frank Paulsen, Daniel Zips, Thais Roque, Nikos Paragios, and Daniela Thorwarth. 2023. Automatic AI-based contouring of prostate MRI for online adaptive radiotherapy. *Zeitschrift für Medizinische Physik* (2023). https://doi.org/10.1016/j.zemedi.2023.05.001

[44] Anna Newcomb, Amber W Trickey, Elena Lita, and Jonathan Dort. 2018. Evaluating surgical residents' patient-centered communication skills: practical alternatives to the "apprenticeship model". *Journal of surgical education* 75, 3 (2018), 613–621.

[45] Vicki L O'Day, Daniel G Bobrow, and Mark Shirley. 1996. The social-technical design circle. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*. 160–169.

[46] Pamela Oliver, Gerald Marwell, and Ruy Teixeira. 1985. A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. *American journal of Sociology* 91, 3 (1985), 522–556.

[47] Wanda J Orlikowski. 1992. The duality of technology: Rethinking the concept of technology in organizations. *Organization science* 3, 3 (1992), 398–427.

[48] Lester J Peters, Brian O'Sullivan, Jordi Giralt, Thomas J Fitzgerald, Andy Trotti, Jacques Bernier, Jean Bourhis, Kally Yuen, Richard Fisher, and Danny Rischin. 2010. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *Journal of clinical oncology* 28, 18 (2010), 2996–3001.

[49] Dong Joo Rhee, Anuja Jhingran, Bastien Rigaud, Tucker Netherton, Carlos E Cardenas, Lifei Zhang, Sastry Vedam, Stephen Kry, Kristy K Brock, William Shaw, et al. 2020. Automatic contouring system for cervical cancer using convolutional neural networks. *Medical physics* 47, 11 (2020), 5648–5658.

[50] Stephanie Santosa and Daniel Wigdor. 2013. A field study of multi-device workflows in distributed workspaces. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 63–72.

[51] Katie A Siek, Yvonne Rogers, and Kay H Connelly. 2005. Fat finger worries: how older and younger users physically interact with PDAs. In *IFIP Conference on Human-Computer Interaction*. Springer, 267–280.

[52] Ashok Sivaji, Azween Abdullah, and Alan Giffin Downe. 2011. Usability testing methodology: Effectiveness of heuristic evaluation in E-government website development. In *2011 fifth Asia modelling symposium*. IEEE, 68–72.

[53] Maaike Van Den Haak, Menno De Jong, and Peter Jan Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology* 22, 5 (2003), 339–351.

[54] MW Van Someren, YF Barnard, and JAC Sandberg. 1994. The think aloud method: a practical approach to modelling cognitive. *London: AcademicPress* (1994).

[55] Catherine G Wolf. 1992. A comparative study of gestural, keyboard, and mouse interfaces. *Behaviour & information technology* 11, 1 (1992), 13–23.

[56] Evan J Wuthrick, Qiang Zhang, Mitchell Machtay, David I Rosenthal, Phuc Felix Nguyen-Tan, André Fortin, Craig L Silverman, Adam Raben, Harold E Kim, Eric M Horwitz, et al. 2015. Institutional clinical trial accrual volume and survival of patients with head and neck cancer. *Journal of Clinical Oncology* 33, 2 (2015), 156.

[57] Haijun Xia, Tovi Grossman, and George Fitzmaurice. 2015. NanoStylus: Enhancing input on ultra-small displays with a finger-mounted stylus. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 447–456.

[58] Matin Yarmand, Chen Chen, Kexin Cheng, James Murphy, and Nadir Weibel. 2024. "I'd be watching him contour till 10 o'clock at night": Understanding Tensions between Teaching Methods and Learning Needs in Healthcare Apprenticeship. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3613904.3642453

[59] Matin Yarmand, Michael Sherer, Chen Chen, Larry Hernandez, Nadir Weibel, and James D. Murphy. 2022. Evaluating Accuracy, Completion Time and Usability of Everyday Touch Devices for Contouring. *International Journal of Radiation Oncology, Biology, Physics* 114, 3 (2022), S96. https://doi.org/10.1016/j.ijrobp.2022.07.515

[60] Matin Yarmand, Borui Wang, Chen Chen, Michael Sherer, Larry Hernandez, James Murphy, and Nadir Weibel. 2023. Design and Development of a Training and Immediate Feedback Tool to Support Healthcare Apprenticeship. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7. https://doi.org/10.1145/3544549.3585894

[61] Stanislaw Zabramski. 2011. Careless touch: A comparative evaluation of mouse, pen, and touch input in shape tracing task. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference*. 329–332.

[62] Stanislaw Zabramski, Suman Shrestha, and Wolfgang Stuerzlinger. 2013. Easy vs. tricky: The shape effect in tracing, selecting, and steering with mouse, stylus, and touch. In *Proceedings of International Conference on Making Sense of Converging Media*. 99–103.

[63] Stanislaw Zabramski and Wolfgang Stuerzlinger. 2012. The effect of shape properties on ad-hoc shape replication with mouse, pen, and touch input. In *Proceeding of the 16th International Academic MindTrek Conference*. 275–278.