Auditory Display for Image-Guided Medical Interventions

by

David Black

A doctoral thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science

Approved Dissertation Committee:
Prof. Dr. Horst Hahn\textsuperscript{1,2} (Chair)
Prof. Dr. Jürgen Schönwälder\textsuperscript{1}
Prof. Dr. Andreas Nüchter\textsuperscript{3}
Prof. Dr. Jörn Loviscach\textsuperscript{4}

\textsuperscript{1} Jacobs University Bremen
\textsuperscript{2} Fraunhofer MEVIS, Bremen
\textsuperscript{3} Julius-Maximilians-Universität Würzburg
\textsuperscript{4} Fachhochschule Bielefeld

Date of Defense: February 11, 2019

Computer Science and Electrical Engineering
Statutory Declaration

<table>
<thead>
<tr>
<th>Family Name, Given Name</th>
<th>Black, David</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matriculation Number</td>
<td>20330212</td>
</tr>
<tr>
<td>Type of Thesis</td>
<td>PhD Thesis</td>
</tr>
</tbody>
</table>

**English: Declaration of Authorship**  I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparisons with future theses. This does not grant any rights of reproduction and usage, however. The Thesis has been written independently and has not been submitted at any other university for the conferral of a PhD degree; neither has the thesis been previously published in full.


Signed:

__________________________________________

Date:

__________________________________________
Abstract

This dissertation provides contributions to the field of auditory display for image-guided interventions and sterile interaction in the operating room. Image-guidance systems for medicines provide clinicians with means of viewing information about a plan that a clinician should perform. These systems help the clinician view patient images acquired before and during an intervention, see the location of tracked tools with relation to images of the human body, or become aware of important factors such as instrument measurement values. Although this information is essential when completing many modern procedures, current methods for delivering intraoperative information rely primarily on computer screens within the operating room to display information. This is a significant drawback, because clinicians must often change between viewing patient and screen and might not be aware of immediate changes, resulting in increased time, inaccuracy, and frustration.

The primary contribution of this dissertation is an investigation into the use of novel auditory display to improve information display in the operating room for image-guided interventions. Novel techniques for transmitting the position of tracked instruments in image-guided procedures are presented, including medical needle placement, robotic endoscope and manual laparoscope guidance, and dissector guidance during open liver resection. These grant the operator auditory display to improve guidance of tracked tools within the body either as an augmentation to existing visual display or to allow screen-free navigation guidance. In addition, a method for hearing fluorescence intensities measured during open brain surgery is presented. Finally, two auditory display approaches are investigated as feedback for novel touchless input using eye tracking and freehand gestures to improve usability and performance.

The results of clinically oriented evaluations presented in this work show that auditory display provides significant advantages in the scope of the selected clinical use cases. The primary benefits of auditory or combined audiovisual display compared to visual-only display include, in most use cases, increases in accuracy, decreases in cognitive workload, and increases in usability. Drawbacks included increases in task completion time and in some cases, increase in cognitive workload. This dissertation clearly shows that auditory display has a promising future in the operating room and warrants in-depth clinical evaluation.
Acknowledgements

I first thank my advisor Horst Hahn for his many years of guidance, troubleshooting, brainstorming, and motivation to help me finish this dissertation, but also for his openness, honesty, and overall positive attitude to my work. Without his guidance during the many projects that made up this dissertation, I would not have made the progress, both personally and academically, that I have enjoyed. Even during his most busiest of days while running the show at Fraunhofer MEVIS, he always took time to discuss any issues on my mind. Furthermore, I would like to thank both Jürgen Schönwälder and Andreas Nüchter for their time and effort reviewing and critiquing this work and being members of the dissertation committee. Much gratitude goes to Jacobs University, who supported me with a stipend during the first years of my study and gave me the freedom to tackle this research. Special thanks to Jörn Loviscach, who made such an outstanding impression during at our first meeting in 2007 that I moved to Bremen to begin my master’s studies. He motivated us to publish as much as we could, even helping send us to conferences to get a taste of real academia. As a member of the dissertation committee, I thank him once again.

I am lucky to have worked intensively with Ron Kikinis, who spent countless hours guiding me over the past years both at Fraunhofer MEVIS and the University of Bremen. He has always been an honest mentor, both professionally and personally. Without him, many of the connections I made with other institutes and researchers would not have been possible. Also, this dissertation would not have been possible without Matthias Günther, who hired me as part of his research staff at the University of Bremen. Thanks to generous travel funding, I was able to partake in research trips and conferences. Thanks to Gracia Kranz for making my time at the University of Bremen delightful. Christian Hansen deserves a great amount of gratitude for his dependable mentorship throughout my career. He helped a ‘music guy’ feel welcome and start to make an impact at an institute for digital medicine. Even after leaving Bremen, he has coauthored numerous articles with me and somehow makes everything seem easy: “You know, David, if you just work on that article/grant/dissertation really hard for a week, you’ll be done with it, then you can submit it!”

My colleagues at Fraunhofer MEVIS are truly the friendliest coworkers anyone could ask for. Thanks to my mentors Tobias Preusser and Christian Rieder for helping plan the steps to success and motivating me to go further. I owe a debt of gratitude to Bianka Hofmann, who has given me so many chances to work on some amazing science and art projects and who has always had an open ear. Merci to Shery Kraß, Gracias to Diego Barrios Romero,
and Dankeschön to Sabrina Haase for being great officemates. Thanks to Torben Pätz for allowing me to work on the SAFE project. To my fellow doctoral students Bojan Kocev and Jan Strehlow, it has been great going through this together and I wish you luck in your final phase. Thanks to Ole Schwen, who is able to distinguish an ‘em dash’ from an ‘en dash’ a meter away and thoroughly reviewed this manuscript. To all MEVIS office staff and admins, I wouldn’t have been able to do it without your tireless help. Thank you and good luck to my mentored students Julia Schütze, Ewald Schmidt, and Kevin Fandio.

I am very honored to have worked in partnerships with a number of excellent scientists to complete this dissertation: Jennifer Nitsch, Hans Meine, Anke Reinschluessel, Mark Herrlich, Rainer Malaka, and Tim Ziemer at the University of Bremen; Maria Luz, Julian Hettig, and Bastian Ganze at Otto von Guericke University in Magdeburg; Michael Unger, Bernhard Glaser, and Thomas Neumuth at ICCAS in Leipzig; Nele Fischer at HWTK Leipzig; Sven Lilge, Jessica Burgner-Kahrs, and Carolin Fellmann at Leibniz University Hannover; Neda Haj-Hosseini at Linköping University; Jayender Jagadeesan at Harvard Medical School; and Rocío López Velazco and Javier Pascau at Charles III University in Madrid. And of course, huge thanks to Arya Nabavi at the International Neuroscience Institute for all the inspiring chats.

To my friends who have been around the last few years, including Ioannis ‘Mr. Südbalkon’ Balas, Hanna Blonska, Thibaud Bucquet, Vanesa Cortes Rodriguez, Thiemo Ende, Andre Hoffmann, Katharina Kessler, Karina ‘Dr. Evil’ Lloyd, Benoit Meriaux, and Leonardo Moura de Araújo: You guys are great, and now you can all stop asking me how my dissertation is going!

To my parents Linda and Larry Black, you have supported me the most of all and I will forever be grateful for everything you have done. To my sisters Sarah and Laurel, thanks for all of your laughs and smiles: I feel privileged to be related to such outstanding young ladies. To Angela Reichert, Roland Diegelmann, Dana Reichert, and Jörg Reichert, you have made me feel so welcome into your families these past years.

Finally, this dissertation would be nothing without the love, support, and amazing cooking of meiner allerliebsten Juliane. You’ve been with me through everything and I love you so much!
# Contents

Statutory Declaration ........................................ iii  
Abstract ................................................................... v  
Acknowledgements ................................................. vii  
1 Introduction ..................................................... 1  
2 Literature Review ............................................... 7  
3 Medical Needle Placement .................................... 27  
4 Telerobotic Transnasal Surgery ............................. 43  
5 Novel Mixed Reality for Laparoscopy ..................... 65  
6 Resection Guidance for Open Liver Surgery .......... 77  
7 Fluorescence-Guided Open Brain Tumor Surgery .... 93  
8 Sterile Interaction Using Eye Tracking ................. 113  
9 Freehand Gesture Interaction ............................. 131  
10 Conclusion and Outlook .................................... 143  
Author’s Publications .......................................... 149  
List of Figures .................................................... 153  
List of Tables ..................................................... 155  
Bibliography ...................................................... 157
Chapter 1

Introduction

Can sound be used to help surgeons and other medical professionals when performing complex tasks, such as instrument placement, or when interacting with operating room systems? This is the central question to be investigated in this thesis, which explores the use of auditory display, a type of sound feedback, to augment existing intervention systems with the aim of giving clinical users enhanced interaction opportunities. The elevator pitch for all of this could be ‘car parking sensors for the surgeon,’ which might seem implausible but accurately describes the essence of much of the work. This thesis attempts to bridge the fields of auditory display and medicine that might seem distantly related, but upon closer inspection, show great promise for integration.
1.1 Medical Background and Challenges

Procedures in medical interventions are currently often supported with image-guidance technology. These methods help enable accurate placement of instruments with relation to the human body using a variety of tracking methods, such as stereotactic or electromagnetic tracking [37]. Typical methods use graphical user interfaces on computer screens in the operating room to information relevant to the procedure, including preoperatively and intraoperatively acquired 2D and 3D images of the patient as well as interventional plans that the clinician should follow. Using this information, the clinician can view the position of interventional tools relative to the patient anatomy to help increase orientation and become aware of possible risk structures near the instrument. This promotes effective and precise completion of the targeted surgical strategy.

Although these hitherto strictly visual methods provide clinicians with valuable tools during planning and intervention, purely graphic representations in image-guided interventions unfortunately exhibit considerable deficiencies. These displays are often overloaded and complex, hindering quick and reliable access to planning information. Viewing angles and monitor placement in the operating room require the clinician to remove concentration from the patient in order to obtain any information from the navigation system. It is difficult to be aware of sudden changes during an intervention, as any information shown on a screen must be seen by the clinician or relayed by an assistant.

In addition to the difficulties faced when navigating tracked instruments, clinicians and other medical assistants, such as scrub nurses, are tasked with interacting with an increasing number of computer systems within the operating room. Despite expanding the range of possibilities for the user, the sterile environment often limits optimal interaction with traditional devices such as keyboards, joysticks, and mice. These devices are usually covered in plastic sheaths that can become quickly soiled during an operation [128], often leading to costly delays [140]. Novel methods for touchless interaction have become increasingly popular [106] with the hope of providing sterile interaction. Common approaches use motion-detection devices [10] or eyetracking [34] for interaction. However, these touchless interaction paradigms lack the tactile feedback found in manual input devices. Unlike keyboard, mice, and joysticks, touchless interaction concepts only provide primary visual feedback, in that the user can only see the result of an initiated action on a screen. Thus, they fail to provide the user with a secondary, tactile means of feedback to inform the user that an action
has been performed. This is a hindrance for implementation in the operating room, where the availability of real-time feedback for interaction has been shown to reduce mental load and improve efficiency and user satisfaction [82].

1.2 Contributions

The main contribution of this work is the investigation and evaluation of so-called auditory display to tackle the difficulties in the operating room concerning image-guided interventions as well as touchless interaction for interventional systems. Auditory display has been defined as “any display that uses sound to communicate information” [152]. Whereas purely visual displays are ubiquitous, auditory displays are found in technical domains ranging from the rather complex environment of an airplane cockpit to everyday applications such as telephone bells, microwave beeps, and automobile parking assistance tones. Sounds produced by these devices provide awareness without the user having to constantly visually monitor an object’s process or status. Thus, auditory displays and their utilization in image-guided interventions could provide outstanding benefits during planning and intervention.

This thesis describes efforts to conceptualize, implement, and experimentally validate the use and effects of auditory and hybrid audiovisual displays for complex medical interventions using modern image-guidance technology. The scope of the works explored in this thesis cover a wide range of medical interventions and applications. For image-guided procedures, these include supporting needle placement (ch. 3), robotic endoscope (ch. 4) and manual laparoscope guidance (ch. 5), dissector guidance during open liver resection (ch. 6), and fluorescence intensity measurement during open brain surgery (ch. 7). For touchless interaction, these include supporting eye tracking (ch. 8) and freehand gesture recognition (ch. 9). Augmenting interventional systems with auditory display is proposed in this thesis to transmit surgical information, such as the distance of an instrument to a risk structure or the level of risk in a certain area, provide immediate feedback for continuous processes, and alert when to take specific action to promote keeping visual attention on the patient. Auditory display is also investigated to enhance interaction with touchless eye-tracking and freehand gesture input systems to increase usability and thus enhance applicability for implementation in sterile environments. This thesis excludes approaches for auditory display as a means of continuous interventional process monitoring,
such as threshold-based warning alarms and auditory display used in anesthesia. Although important for the success of an intervention, sound in the context of continuous process monitoring is not used for the direct task of image-guided navigation. For detailed information on auditory display for anesthesia monitoring, see Sanderson et al. [129].

The paths used to investigate the methods described in this thesis are similar. All projects involve feature a user-centered approach including a contextual inquiry that feature intense discussions with potential users, especially medical professionals at partner clinics. After defining clinical needs, first methods of auditory display for each application were investigated iteratively with both clinicians and novice users. Finally, all developed methods were evaluated in user studies to determine benefits of the approaches with respect to both quantitative and qualitative measures. Quantitative measures included (depending on the clinical application) task completion time, placement accuracy, time spent viewing screens, path-following efficiency, verbal response accuracy, and reaction time. Primary qualitative measures included responses to both the NASA-TLX assessment tool for perceived workload [69] as well as the van der Laan scale for system acceptance in terms of usability and satisfaction [94]. This thesis does not focus on the general design processes and choices for auditory display. This has already been covered in depth in a previous work [11], which includes a review of the auditory environment in the operating room, auditory display design methods and considerations, and preliminary evaluations. In addition, my own previous publications [12, 14, 17, 19, 20] detail further design considerations and techniques and describe user evaluations for auditory display for relevant clinical settings and applications. Knowledge gained from these auditory display design processes and considerations, thus, make up part of the preliminary work necessary to develop the methods presented in this thesis.

In addition to research performed at Jacobs University, Fraunhofer MEVIS, and the University of Bremen, the methods proposed in this thesis have been developed in close collaboration with a number of partner institutes, university groups, and clinics. These include Otto-von-Guericke University in Magdeburg, Germany (chs. 2, 3, and 9), the Laboratory for Continuum Robotics at Leibniz University in Hanover, Germany (ch. 4), Brigham and Women’s Hospital and Harvard Medical School in Boston, USA (ch. 5), Linköping University, Sweden (ch. 7), the Innovation Center Computer Assisted Surgery (ICCAS) in Leipzig, Germany (ch. 8), the International Neuroscience Institute in Hanover, Germany (chs. 2 and 4), Robert-Bosch Hospital in Stuttgart, Germany (ch. 6), and Asklepios Clinic Barmbek in Hamburg, Germany (ch. 6).
This thesis contributes a substantial set of new scientific knowledge to the growing field of auditory display for image-guided interventions. Several of the projects described are the first investigations of auditory display in each of their respective medical domains. In addition, the projects provide a blueprint for successful integration of researchers from the interdisciplinary fields of auditory display design, medical engineering, and medicine, the three of which have seldom been combined for fruitful cooperation. Although the evaluations of the methods were performed in laboratory settings, all included a clinical orientation and featured valuable input and guidance from partner clinicians. The result of these partnerships is a small but significant set of publications that has already set the stage for the growing field of auditory display in medicine.

1.3 Structure of the thesis

This thesis is divided into chapters based on publications that correspond to the range of applications for which auditory displays have been researched and evaluated. After a review of related literature, chapters 3 to 6 showcase auditory display for instrument guidance, chapter 7 details auditory display for fluorescence measurement, and chapters 8 and 9 present auditory display for sterile interaction methods for operating environments.

- **Chapter 2** presents an overview of the state of the art of auditory displays that have been described in connection with image-guided interventions, including auditory display methods implemented as well as relevant results. A discussion of motivations, methods and results provides insight into the current situation as well as suggestions for future directions of research.

- **Chapter 3** describes an auditory display to augment or replace visual feedback for navigated needle placement. In contrast to existing auditory display approaches which augment but still require a visual display, this method also allows view-free needle placement. An evaluation compares both auditory and combined audiovisual feedback against typical visual methods.

- **Chapter 4** presents an auditory display to enhance interaction with teleoperated continuum robots for placement tasks for a typical transnasal intervention through the sphenoidal sinus. A test environment simulates a typical transnasal intervention through the sphenoidal sinus using a continuum robot and compares visual and audiovisual displays.
Chapter 1. Introduction

- **Chapter 5** describes the design and validation of a novel mixed reality head-mounted display for intraoperative surgical navigation. As part of a larger project involving mixed reality, an auditory display is described and evaluated using a peg identification and transfer task.

- **Chapter 6** introduces an auditory display system for resection during open liver surgery to support guiding the tracked instrument towards and remaining on a predefined resection line. A combined audiovisual display is evaluated against an existing visual navigation system.

- **Chapter 7** describes an auditory display for transmitting levels of fluorescence intensity values measured by a hand-held probe for use during fluorescence-guided open brain surgery. The auditory display is evaluated using a set of liquid phantoms which mimic those found in actual interventions.

- **Chapter 8** presents an eye-tracking input system with an auditory display using both earcons and parameter-mapping sonification for typical scrub nurse tasks by replacing the lost tactile feedback when using touchless input. An evaluation in a laboratory operating room environment compared auditory display with visual display with respect to reaction time and a series of subjective measures.

- **Chapter 9** explores the use of auditory display to improve free-hand gesture interaction for operating room interaction. Gestures using a Leap motion controller were augmented with auditory icons and continuous, model-based sonification. Three concepts were evaluated using a sphere-selection task and a video frame selection task.

In each chapter, laboratory evaluations with a strong clinical orientation are reported. Finally, **Chapter 10** summarizes the results, challenges, and contributions of each of the previous chapters and provides insight into future work in the field of auditory display for image-guided medicine and sterile interaction. Publications have been edited for clarity, consistency, and minor errors. For all chapters, I provided the primary or complete contribution to auditory display design and programming. Chapter 2 was primarily researched and composed by me. For all remaining chapters, I provided major contributions to contextual inquiry, overall system specification and design, preliminary testing, experimental design and execution, and article composition. For chapter 5, I provided supplemental contributions to these aspects. For chapter 3 to 9, I provided minor contributions to statistical analysis. In addition, all image-guidance or tracking systems upon which the auditory display were based were developed or configured by partner institutions.
Chapter 2

Literature Review


About this chapter

Auditory display is a growing field that has been largely neglected in research in image-guided interventions. Despite benefits of auditory displays reported in both the literature, adoption in medicine has been slow. This chapter provides an overview of the current state of the art of the use of auditory display in image-guided medical interventions, including methods for avoidance of risk structures and instrument placement and manipulation. The chapter discusses challenges that face researchers wishing to develop more meaningful auditory display designs and evaluates the benefits and drawbacks of auditory display in image-guided medicine.
2.1 Auditory Display in Image-Guided Therapy

Modern medical image-guided interventions depend on reliable access to patient data to ensure a successful procedure. Navigated interventions typically employ virtual images of planning data overlaid on images of patient anatomy to aid surgeons during the procedure, for instance, to view the location of a tracked instrument in relation to patient anatomy, to locate the target site, or to become better aware of the locations of risk structures or objects of interest. The field of image-guided interventions has grown greatly over the last 20 years thanks to progress in medical imaging and computing technology. For an overview of image-guided intervention technology and clinical applications, see Cleary and Peters [37]. Using image guidance during an intervention, clinicians can access important information that was previously unavailable, typically on a computer screen placed in the operating room.

However, despite the benefits of image guidance in medicine, displaying information on a screen is sometimes not ideal [106] and alternatives to traditional computer screens are being researched. Clinicians must remove their view from the operating situs to receive information [67], meaning important notifications might not be perceived [157], and 3D information might not be correctly interpreted [22]. To remedy some of the deficiencies inherent in visual display, the relatively new field of auditory display [92] presents an interesting possibility for image-guided medicine. Auditory display harnesses sound to present information; changes in a data source can be mapped to parameters in a sound synthesizer so that a user can hear information as opposed to viewing it on a screen. This chapter presents a review of the literature on the use of auditory display in image-guided interventions, including systems for volumetric resection, telerobotic suture tying, resection path guidance, needle placement, and temporal bone drilling. Various motivations, auditory display approaches and evaluation results are presented and discussed, and the primary problems and future trends in auditory display in image guidance are presented.

2.2 Literature Search

A search of the literature was performed using a combination of the following search terms: ‘auditory display,’ ‘image guidance,’ ‘sonification,’ ‘image-guided navigation,’ and ‘auditory navigation.’ We performed forward and backward searches using PubMed [146] and Google Scholar [57] for related
literature. Cited, citing, and similar articles that satisfied the eligibility criteria (see below) were thus included in the review. We did not include language restrictions in the search. Literature search and review were performed by two independent reviewers (David Black, Christian Hansen).

2.2.1 Eligibility Criteria

We considered literature for inclusion that included auditory display as an integral part of navigated image-guided intervention support. These included scientific articles that described both laboratory prototypes, clinical applications of auditory display, as well as detailed, published descriptions of systems yet to be evaluated in a laboratory or clinical setting. We excluded literature that focused on sound as a means of continuous interventional process monitoring, such as basic warning alarms auditory display used in anesthesia. Although important for the success of an intervention, sound in the context of continuous process monitoring is not used for the direct task of image-guided navigation. For detailed information on auditory display for monitoring for anesthesia, see Sanderson et al. [129].

2.2.2 Data Extraction

During the search process, we retrieved articles meeting the aforementioned eligibility criteria for further assessment. The following information was extracted from each article:

- Interventional tasks to be supported by auditory display
- Motivations for including auditory display for navigational support
- Auditory display methods employed
- Evaluation designs and findings
- Clinical considerations and discussion specifically concerning the use of auditory display
2.3 Results

The results of the search yielded 15 articles [12, 18, 22, 35, 36, 39, 67, 90, 99, 141, 150, 155–158] that met the eligibility criteria. The eligible articles cover a wide range of interventional tasks, implemented auditory display methods, evaluation styles and environments, and findings.

2.3.1 Interventional Tasks Supported by Auditory Display

The selection of literature reveals a broad spectrum of interventional tasks supported by auditory display, see Tables 2.1 2.2. Four of the 15 articles concern needle placement: Wegner and Karron [155], Wegner [156], Black et al. (2013) [12], Bork et al. [22], and Black et al. (2017) [18]. Specifically, Wegner [156] explores a generalized instrument placement task with a tracked drawing device that is meant to aid “… a procedure requiring an insertion trajectory.” Black et al. [12, 18] explore radiofrequency needle ablation targeting lesions, and Bork et al. [22] support needle biopsy targeting lesions.

Five of the articles support temporal bone drilling applications or mastoidectomy. Cho et al. first support monitoring the “distance between the drill tip and important organs,” [36] and later [35] support monitoring the distance between the drill tip and facial nerve, as well as the distance between scala vestibuli, and scala tympani (parts of the ear). Voormolen et al. [150] support monitoring the distance between the drill tip and the facial nerve and sigmoid sinus. Dixon et al. [39] implement auditory display alerts for monitoring the distance of the drill to dura and carotid arteries in skull base surgery. Luz et al. [99] support image-guided mastoidectomy.

Three of the articles describe aids for tissue resection. Willems et al. [157] and Woerdeman et al. [158] support volumetric lesion resection for neuronavigation. Hansen et al. [67] develop an auditory display for open liver surgery for transferring a preoperatively planned initial resection path onto the surface of the liver.

Finally, Kitagawa et al. [90] implement auditory display as a means of sensory substitution for the loss of haptic feedback encountered when performing suture manipulation during telerobotic surgery. Strauss et al. [141] support functional endoscopic surgery of the paranasal sinuses with a collision warning system for monitoring distance of the surgical instrument to the frontal skull base, lamina papyracea, and internal carotid artery.

---

1 Authors were involved in [12, 18, 67]
### Table 2.1: Review of clinical applications and auditory display approaches using proximity alerts

<table>
<thead>
<tr>
<th>Clinical Application</th>
<th>Proximity AD Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kitagawa et al. (2005) [90]</td>
<td>Sensory substitution for forces during robotic suture ties</td>
</tr>
<tr>
<td>Strauss et al. (2010) [141]</td>
<td>Endo- and transnasal surgery</td>
</tr>
<tr>
<td>Voormolen et al. (2012) [150]</td>
<td>Temporal bone drilling for target access during neuronavigation</td>
</tr>
<tr>
<td>Cho et al. (2013) [36]</td>
<td>Protecting facial nerve during otologic surgery by monitoring safe region</td>
</tr>
<tr>
<td>Cho et al. (2014) [35]</td>
<td>Guiding cochlear implantation</td>
</tr>
<tr>
<td>Luz et al. (2015) [99]</td>
<td>Mastoidectomy using distance control and navigated control concepts</td>
</tr>
</tbody>
</table>

### Table 2.2: Review of clinical applications and auditory display approaches using continuous auditory display

<table>
<thead>
<tr>
<th>Clinical Application</th>
<th>Continuous AD Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wegner et al. (1997-8) [155, 156]</td>
<td>Generalized 3D medical instrument placement</td>
</tr>
<tr>
<td>Hansen et al. (2013) [67]</td>
<td>Path marking in open liver resection with surgical ultrasound dissector</td>
</tr>
<tr>
<td>Black et al. (2013) [12]</td>
<td>Radiofrequency ablation needle guidance</td>
</tr>
<tr>
<td>Bork et al. (2015) [22]</td>
<td>Needle biopsy to target virtual lesions</td>
</tr>
<tr>
<td>Black et al. (2017) [18]</td>
<td>Radiofrequency ablation needle guidance</td>
</tr>
</tbody>
</table>
2.3.2 Clinical Motivations for Exploring Auditory Display

The motivations for developing auditory display to aid image-guided interventions arise from shortcomings in traditional image-guided intervention methods described in the literature. Three primary motivations named in the literature include:

- increasing awareness of structures surrounding the tracked instrument
- reducing attention to the screen or increasing attention to the patient or test phantom
- helping clinicians correctly interpret (multidimensional) navigation data.

Aiming to improve clinician interaction with visual displays and change view behavior was also cited in much of the literature [12, 18, 35, 36, 67, 141, 150, 155, 156]. Investigators commented on the necessity of clinicians to draw attention away from the situs in order to view the navigation screen. Willems et al. [157] argue that “to appreciate the visual information offered . . . the surgeons’ attention (the visual focus) will need to be drawn away from the actual surgery. This will result in the images being used only at intervals chosen by the surgeon.” Hansen et al. [67] state that “the navigation system needs to be frequently consulted by surgeons, which leads to increased mental load and time pressure during surgery. The surgeon’s attention to the working area is interrupted by viewing the navigation system’s screen.” Cho et al. [36] note that “when using a navigation system, the surgeon’s visual focus must move between the operating field and the navigation monitor to identify the position of the drill, causing a temporary interruption in the temporal bone dissection.” Wegner [156] cites as motivation for auditory display “users who cannot tolerate the encumbrance of graphical display hardware, and whose visual faculties have pre-existing obligations, such as addressing the task at hand.”

Most articles that focus on threshold alerts [39, 99, 141, 150, 157, 158] understandably mention the aim of increasing awareness of the anatomy or critical structures surrounding the tracked instrument. For instance, Dixon et al. state that “. . .surgery can be technically demanding and requires a continuous appreciation of the surrounding critical structures.” Strauss et al. mention a desire to “improve the situational awareness of the surgeon,” because “in the field of surgical navigation, situational awareness for the described conventional task is not optimal.” [141] Woerdeman et al. lament that

---

2Passages from Strauss et al. [141] translated from the original German into English by author.
“complete spatial awareness at all times can be compromised during [image-guided surgery].” [158] According to Voormolen et al., “standard neuronavigation does not adequately notify the surgeon about where he/she is drilling in relation to surrounding temporal bone critical structures.” [150]

A third primary motivation cited by a number of articles is the correct interpretation of navigation information [22, 141, 157, 158]. Bork et al. [22] cite the lack of usable depth information in augmented reality applications: “when [augmented reality] visualization is implemented as a simple superimposition of virtual objects on the video stream, the virtual objects appear to float above the anatomy. This lack of correct depth perception has been recognized as a major challenge,” hoping that auditory display can ameliorate this lack of depth information in the camera view. Strauss et al. [141] state that “the surgeon is continuously required to translate the supplied information into a 3D model. This approach is laborious and prone to error.” Willems et al. [157] and Woerdeman et al. [158] both note the difficulty in interpreting conventional 2D views of 3D scenarios.

Further motivations include the ability to hear structures occluded in visual display [22], substitute the loss of tactile sense during robotic surgery [90], lessen simulator sickness [156] and vertigo [155], reduce clinician workload [12, 39, 67, 141, 155], and reduce memory burden [156].

### 2.3.3 Methods of Auditory Display

An auditory display is one that uses sound rather than a screen to communicate information [152]. These use data from some source that is typically mapped to changing parameters of a sound generator which generates acoustic output. Auditory displays are quite common in everyday life, with applications including speech, radios, music, alarm clocks, bells, telephone ring tones, microwave buzzers, sirens, and horns [71]. Historically, scientific auditory display has been sparsely employed, for example to bring seismometer recordings into audible range for analysis [138]; the first International Conference on Auditory Display was held in 1992 [92]. Using auditory display for guidance tasks in related fields has been described for applications such as obstacle avoidance and route finding for blind pedestrians, encouraging athletes towards more efficient movements, or aiding patients during rehabilitation [115]. In the case of image-guided interventions, the data source is typically distance information delivered by the navigation system.
In addition to ameliorating the shortcomings in traditional image-guided interventions described above, Wegner and Karron [155, 156] describe multiple benefits of auditory display for interventional use, including the omnidirectionality of audio allowing for information display without line-of-sight, the relatively open auditory perceptual channel, reduced computational demands of audio synthesis, and the ability of humans to perceive parallel streams of audio. Further general benefits of auditory display including improving ergonomics, for instance, by reducing the number of head and neck movements to switch between viewing various displays [25] or to promote rapid detection of events in high-stress environments [92]. Auditory display has been shown to be fairly easy to learn [152], and even engaging or fun to use [131].

Although there are multiple methods of auditory display available to the sound designer, the reviewed literature includes three primary auditory display methods: alerts, auditory icons, and parameter mapping.

Alerts

Alerts are sounds that are played back when the source data reaches a predetermined threshold. These are common in the operating environment [129]. The purpose of alerts is to indicate that an event has taken place or is about to occur, thereby prompting the listener to take action. In the reviewed literature, alerts have been described by 6 of the 14 articles. For instrument tracking, the alert plays back when the distance of the tracked instrument to a certain structure has passed a predetermined distance threshold [35, 36, 99, 141, 150, 157], or in the case of Kitagawa et al. [90], when applied during telerobotic suture tying tension reaches a desired threshold.

For volumetric resection, Willems et al. [157] implement an alert when the tip of the tracked instrument encroached a predefined contour. The alert played back twice per second with a frequency of 510 Hz and duration of 0.1 seconds. For temporal bone drilling, Cho et al. (2013) [36] create three absolute distance margins of 2, 4, and 6 mm from the facial nerve, which correspond to alerts of 900, 600, and 300 Hz, respectively, played back for 20 ms. In a second article [35], two alerts play back when the tip was within 5 mm of either the scala vestibuli or the scala tympani. If the distance to the scala vestibuli is greater than the distance to the scala tympani, an alert of 900 Hz is played; if the distance to the scala tympani is greater than the distance to the scala vestibuli, an alert of 300 Hz is played. Strauss et al. [141] describe playing back an alert when the instrument reaches a “minimal distance,” but detailed descriptions of the auditory display method are not provided in the
2.3. Results

Tracked instrument
Auditory encoding
Planned Path
Risk Structure
Object of Interest

Figure 2.1: Various approaches to map tracking data to auditory display found in the literature. From top to bottom: risk avoidance using safety margins, resection path following, 3D trajectory following, and temporal distance coding.
text. For telerobotic suture tension, Kitagawa et al. [90] describe an “[audio feedback] mode, which provided a single tone when the magnitude of the applied tension reached the [optimal] manual tension,” although detailed descriptions of the auditory display method are not provided.

**Auditory Icons**

Somewhat more complex than alerts are auditory icons, which are “everyday sounds to convey information about events by analogy to everyday sound-producing events.” [30] These icons are used in a similar way to visual icons: they map system events to those found in everyday listening, mimicking such sounds as throwing trash in a bin, commonly employed when deleting a file in a desktop graphical environment. Short auditory icons use the richness of everyday sounds and their ease of comprehension to link sounds to events.

In the case of Dixon et al. [39], aforementioned simple abstract alerts were first developed, but preliminary tests suggested that participants found it “difficult to distinguish acoustically which anatomical structure was close and how far away it was.” Auditory icons representing the dura and carotid arteries were developed to be easier to learn. For instance, the sound of an arterial Doppler trace was used to represent proximity to a major artery. Dixon et al. manually set safety margins to 2 and 3 mm.

**Parameter Mapping Models**

In contrast to alerts and auditory icons, parameter mapping links continuous changes in one set of data to continuous changes in audio parameters, providing a higher level of complexity. In essence, the underlying data delivered by the navigation system are used to ‘play’ a realtime software instrument according to those changes. Because audio has a wide range of parameters [156] that may be altered, such as frequency, intensity, and timbre, continuous parameter mapping is also suitable for displaying multivariate data. This technique attempts to make the listener an active participant in the listening process by providing interactive, changing mappings that relate data to audio. This method is useful for smoothly representing continuous changes in events.
In image-guided intervention support systems which implement parameter-mapping auditory display, the tracked instrument itself is in essence the physical musical instrument: the clinician plays the realtime software instrument by moving the tracked instrument. The range of parameter mappings found in the literature extend from fairly simple frequency and volume mappings [150, 158] up to complex methods such as 3D audio spatialization and wave-terrain synthesis [155, 156].

For volumetric resection for neuronavigation, Woerdeman et al. [158] adapt a previous approach of the group [157], which employed a simple alert, to play back a parameter-mapping alert. They describe a “soft warning sound (an intermittent pure tone)” with a duration of 0.1 seconds that plays back at a rate of 3 times per second at a distance of 5 mm from the tumor outline. Upon entering the 5 mm threshold, volume and tone frequency increased proportionally until the outline of the tumor was reached. After encroaching the tumor outline, a “continuous pure tone” was played back. Voormolen et al. [150] cite and employ this method for temporal bone drilling but do not describe further adaptation details.

Biopsy needle placement support is described by Bork et al. [22], who use a method of parameter mapping called temporal distance coding [53], in which the time an object is rendered depends on the distance from the tracked instrument. In this case, auditory temporal distance coding allowed playback of the distance from the tip of the biopsy needle to objects of interest within the AR environment. Virtual ‘spheres’ propagate from the needle tip at a certain speed. The longer it takes for these spheres to collide with the objects of interest, the longer a metronome sound is played back. Once an object is ‘hit’, a bell tone is played. Thus, the more metronome sounds are played before a bell tone is played, the further the object is from the needle tip. The user is not explicitly guided towards a target using auditory display, but rather receives information concerning the distance of objects of interest in the area.

A different method for needle placement is investigated by Black et al. [12, 18]. This auditory display encodes the distance of the tip of the needle to the correct insertion point, the distance of the shaft to the correct position, and the depth as the distance of the tip to the target lesion. Two auditory display methods are described. In both methods, the task of needle placement is split into tip placement, handle placement, and insertion phases. The first method employs a tone with a moving pitch and a reference tone; the pitch of the moving tone is mapped to the distance in the \( y \)-axis. These are brought together, creating an auditory display that mimics tuning an instrument. Distance in the \( x \)-axis is mapped to the inter-onset interval of train of tones, from 250 ms at the outer edge of the navigation area to 100 ms at the center. The
second method further separates motion along the $x$-axis and $y$-axis. Placement along the $x$-axis is first performed using changes in inter-onset interval, repeated again for movement along the $y$-axis corridor. After correct tip and handle placement, the needle is inserted and depth to target is mapped to the increasing pitch of 10 consecutive tones, after which a bell tone is played back upon reaching the target lesion.

Hansen et al. [67] support resection line marking for open liver surgery with a parameter-mapping auditory display. In this method, the navigation system delivers the nearest distance between the instrument tip and the planned resection line. The distance is divided into three margins: ‘safe,’ ‘warning,’ and ‘outside.’ When the instrument is in the safe margin, signaling to the clinician that the position is correct, a confirmation tone is played back with a frequency of 698.5 Hz and an inter-onset interval of 660 ms at the center of the safe margin and 180 ms at the edge of the safe margin. In the warning margin, the distance is mapped to inter-onset interval, pitch, and tone length. Pitches to the left of the resection line become consecutively lower, while those to the right of the resection line become consecutively higher, thus providing directional information. Outside the warning zone, no sound is played to prevent unwanted sound when the instrument is not near the resection line.

Wegner and Karron [155] map a discrete error function in the plane perpendicular to the trajectory path to MIDI\textsuperscript{3} tones. For placement in this plane, a chordal drone is produced, with deviations from the correct placement producing inharmonicity. Another tone was produced at regular points along the trajectory path to ‘tag’ the distance traveled. A second method [156] employs beat interference between 3 pairs of sinusoids which correspond to 3 axes in space. By reducing the beat interference between each of the 3 pairs, correct position is found.

Figure 2.1 visualizes the primary mapping approaches encountered in the literature.

2.3.4 Experimental Designs and Findings

The variety of experimental designs used to evaluate the reviewed literature ranges from informal evaluations to phantom studies in laboratory conditions to clinical evaluations.

\textsuperscript{3}Musical Instrument Digital Interface, a protocol for electronic musical instrument communication
Three of the 14 reviewed articles provide informal evaluations without statistical data gathering or analysis; see Table 2.3 “Informal Evaluations” for an overview. Wegner and Karron [155] provide solely a technical description of their range of auditory display methods for generalized tracked instrument placement. Wegner [156] states that “informal usability testing” was completed, but does not further elaborate. Black et al. (2013) [12] perform informal, ‘think-aloud’ evaluations [97] of 2 auditory display methods for ablation needle placement with 8 non-expert participants. Comments were gathered during the interviews and suggest general satisfaction with performance during the placement task.

The majority (12) of the reviewed literature describe evaluations in laboratory conditions on phantoms, see Table 2.3. Of these, 4 of 15 also performed a clinical evaluation described in the same article, see Table 2.4 for an overview.

**Table 2.3: Overview of informal and laboratory evaluations in literature**

<table>
<thead>
<tr>
<th>Informal Evaluation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wegner (1998) [156]</td>
<td>Informal usability testing</td>
</tr>
<tr>
<td>Black et al. (2013) [12]</td>
<td>8 participants: talk-aloud walk through and interview</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Laboratory Evaluation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kitagawa et al. (2005) [90]</td>
<td>5 surgeons: suture ties with different materials for manual tying and no-feedback, auditory, visual, and audiovisual displays</td>
</tr>
<tr>
<td>Willems et al. (2005) [157]</td>
<td>5 surgeons: volume resection on floral foam phantoms using auditory display and standard visual navigation.</td>
</tr>
<tr>
<td>Woerdeman et al. (2009) [158]</td>
<td>4 surgeons: volume resection using auditory, conventional display, and heads-up display</td>
</tr>
<tr>
<td>Strauss et al. (2010) [141]</td>
<td>5 surgeons: reported and actual distance to structures with using conventional and combined audiovisual display</td>
</tr>
<tr>
<td>Voormolen et al. (2012) [150]</td>
<td>5 surgeons: bone drilling in phantoms, comparing combined audiovisual display and conventional display</td>
</tr>
<tr>
<td>Cho et al. (2013) [36]</td>
<td>1 surgeon: bone drilling with and without audiovisual display</td>
</tr>
</tbody>
</table>
**Chapter 2. Literature Review**

<table>
<thead>
<tr>
<th>Clinical Evaluation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hansen et al. (2013) [67]</strong></td>
<td>12 surgeons: resection line marking with audiovisual and conventional navigation display. Audiovisual display reduced the percent of time viewing the screen, increased accuracy of the marking task, and increased task completion time.</td>
</tr>
<tr>
<td><strong>Dixon et al. (2014) [39]</strong></td>
<td>7 surgeons: dissection and clivus ablation with and without audiovisual display. Using the audiovisual display reduced perceived workload scores for mental demand, effort, and frustration.</td>
</tr>
<tr>
<td><strong>Luz et al. (2015) [99]</strong></td>
<td>18 surgeons performing mastoidectomy manually, with navigated control, and distance control. Distance control resulted in lower task completion time and workload.</td>
</tr>
<tr>
<td><strong>Bork et al. (2015) [22]</strong></td>
<td>15 participants: lesion targeting with simple overlay, auditory feedback, visual feedback, and audiovisual feedback. Audiovisual feedback resulted in most target hits and least localization error. Auditory, visual, and audiovisual more accurate, slower than simple overlay. Audiovisual display outperformed auditory and visual display in accuracy, task completion time, and number of lesions hit.</td>
</tr>
<tr>
<td><strong>Black et al. (2017) [22]</strong></td>
<td>12 participants: needle tip and shaft placement with auditory feedback, visual feedback, and audiovisual feedback. Audiovisual feedback resulted in reduced localization error. Purely auditory display exhibited longer task times, combined display was fastest.</td>
</tr>
</tbody>
</table>

**TABLE 2.4: Overview of clinical evaluations in literature**

<table>
<thead>
<tr>
<th>Clinical Evaluation</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Woerdeman et al. (2009) [158]</strong></td>
<td>1 surgeon: 6 patients resection with auditory display randomly activated. No specific effect of auditory display on instrument tip speed. Subjective reports of improvements in decision-making.</td>
</tr>
<tr>
<td><strong>Strauss et al. (2010) [141]</strong></td>
<td>4 surgeons: functional endoscopic sinus surgery. Complication rate reduced and preparation time increased when using audiovisual display.</td>
</tr>
<tr>
<td><strong>Cho et al. (2013) [36]</strong></td>
<td>1 surgeon: 2 cochlear implantations, 1 acoustic tumor resection. Warning margins with auditory display allowed drilling continuously without removing view from situs.</td>
</tr>
<tr>
<td><strong>Cho et al. (2014) [35]</strong></td>
<td>1 surgeon: 2 cochlear implantations. Auditory display helped locate correct cochleostomy point while keeping focus on microscope.</td>
</tr>
</tbody>
</table>

For the task of telerobotic suture tying reported by Kitagawa et al. [90], 5 surgeons completed suture ties using no feedback, auditory feedback, visual feedback, and audiovisual feedback after 1 hour training with the robot system. Findings indicate that consistency of tying tension with sensory substitution using visual and audiovisual displays were superior to those of hand ties, and that the consistency of tying tension using auditory display were comparable to those of hand ties.
2.3. Results

Willems et al. [157] compare volume resection on floral foam phantoms with 3 experienced surgeons who each completed one resection each using both auditory display and standard visual navigation. Results indicate that using auditory display, the similarity of the resected volume to target volume increased marginally, the amount of target tissue not removed was reduced, but that the amount of non-target tissue removed increased.

Using a similar task, Woerdeman et al. [158] describe an evaluation with 4 surgeons performing volume resection with auditory display, conventional IGS display, and a heads-up display. Task completion time between auditory display, conventional display and heads-up display did not differ, and target volume removal did not differ between auditory and conventional displays. However, auditory display was subjectively perceived to improve performance compared to conventional display.

Strauss et al. [141] compared surgeon-reported and actual distance measurement points from instrument to risk structures with 5 ‘advanced beginner’ surgeons using a combined audiovisual collision warning system. Results indicate that the audiovisual display improved reported accuracy 76% over conventional display.

Voormolen et al. [150] evaluate 5 surgeons each performing a temporal bone drilling task in two phantoms, once with conventional image guidance and once with the combined audiovisual assistance. Using the audiovisual system, no critical structures were damaged (opposed to three structures using conventional methods), and participants reported improved subjective orientation and improved tumor exposure with the system.

Cho et al. (2013) [36] describe a laboratory study with one inexperienced surgeon who drilled 10 bone phantoms, 5 using an audiovisual display and 5 without navigation. The drill distance to the facial nerve was recorded to determine when the surgeon encroached the safe margin of 2 mm to the nerve and when the nerve was hit. Using no navigation, the nerve was hit in 4 of 5 attempts, whereas with audiovisual display the nerve was hit once. In addition, the uniformity of the safe margin in the resected area appeared greater with audiovisual display.

Hansen et al. [67] compare resection line marking on a floral foam phantom with 12 surgeons using combined audiovisual and conventional 3D navigation display. Findings indicate that the auditory display reduced the percent of time viewing the screen from 96% using visual display to 10% using combined audiovisual display. Auditory display increased accuracy of the marking task, but also increased task completion time.
Dixon et al. [39] report an evaluation of 14 cadaver specimens with 7 surgeons who each performed dissection and clivus ablation on 2 heads, once each using conventional display and audiovisual display. Using the audiovisual display reduced NASA-TLX perceived workload [69] scores for mental demand, effort, and frustration.

Luz et al. [99] evaluate 18 surgeons who performed a simulated mastoidectomy surgery using navigated control and distance control, as well as manually without image guidance. Effects on surgical performance, physiological effort, workload and situation awareness were evaluated. Using simpler distance control instead of more complex navigated control led to faster task completion and lower subjective workload.

Bork et al. [22] evaluate lesion targeting using a biopsy needle with 15 non-clinical participants. Each participant completed 3 attempts using simple lesion overlay, auditory-only feedback, visual feedback, and audiovisual feedback. Participants verbally confirmed reaching each lesion point. Results show that targeting using combined audiovisual feedback resulted in most target hits and least localization error. Auditory, visual, and audiovisual displays improve accuracy but resulted in slower task completion times than the simple overlay. Audiovisual display outperformed auditory and visual display with respect to accuracy, task completion time, and number of lesions hit.

Black et al. (2017) [18] evaluate tip and handle placement of an ablation needle with 12 non-clinical participants. Each participant completed tip and needle placement using auditory, visual, and combined audiovisual display. For each placement, the participants verbally confirmed when tip and needle were placed to their satisfaction. Results showed that combined audiovisual display provided the highest accuracy of the three feedback modalities. Purely auditory display exhibited longer task completion times and increased subjective workload compared to visual and combined audiovisual display.

Clinical Evaluations

Four of the 15 reviewed articles evaluated their approaches in clinical conditions in addition to the laboratory studies, see Table 2.4.
2.4 Discussion

This review presents, to the authors’ knowledge, the first overview of the state of the art of auditory display for image-guided interventions. The articles included in the review cover a range of interventional tasks to be supported, auditory display methods, and evaluation designs and findings. Whereas a majority of the literature covers the use of auditory display to inform the clinician of risk structures, thereby prompting the clinician to navigate away from a certain object of interest, a number of articles attempt to aid clinicians in navigation towards a target itself.

The body of results of the reviewed literature show that in most cases, systems with auditory display were found to be beneficial. Advantages include improved recognition of the presence of or distance to anatomical risk structures [36, 141, 150], reduced complication rate [141], improved
placement accuracy [18, 22, 67, 141], improved resection volume similarity [157], improved orientation [150], reduced workload demands [39]. Reported drawbacks included increased task time [18, 67, 141], increased workload [18] and increased amount of non-target tissue removed during volumetric resection [157].

Considering the wide range of tasks that are currently supported with image guidance [37] and primarily positive evaluation results of the use of auditory display in the reviewed literature, it is surprising that only a limited amount of investigations have attempted tackling auditory display. Even when auditory display has been integrated into support for image guidance, the majority of attempts usually only implement a simple threshold-based alert. The paucity of investigations into more complex navigational aids might be traced to an unfamiliarity with the relatively nascent field of auditory display and its possibilities in enhancing guidance tasks, possibly prompted by clinician dissatisfaction with previous experiences with alarms. The abundance of other sounds in the operating environment, including speech and instrument noises, plays a role in the distrust or rejection of new auditory displays. This is discussed in the reviewed literature: an editorial comment in response to Willems et al. [157] recognizes the benefit of reducing the necessity of viewing a navigation screen, but is pessimistic of its clinical application:

“We doubt whether we would like to have such an auditory warning system in the operating room creating distracting sounds. Especially when surgery comes to critical areas, a beeping neuronavigation system may be annoying, since the operating room is already filled with acoustic warning systems of the anesthesiologist, with which an additional system should also not interfere. We should seek ways to increase the comfort for the surgeon in the operating room, allowing concentration on the surgical field, supported by enhanced guidance systems using modern 3D visualizing techniques.”

This sentiment is reflected in statements by authors of the reviewed literature, who caution that auditory display methods should take the sounds in the existing operating room into account during the design phase to reduce annoyance or overburden the environment [22, 36, 39, 67, 90, 155, 157]. To be sure, the operating room is a noisy environment, with average sound levels considerably higher than those of other workplaces [107]. However, according to Katz et al., “there is little evidence to demonstrate a direct association between excessive operating room noise and poor surgical outcomes.” [84]
Music in the operating room has become commonplace [107], and surgeon-selected music in the operating room has been reported to enhance performance [3] and reduce workload [54]. In addition, Moorthy et al. [109] report that surgeons can effectively block out unwanted noise in the operating room.

As a whole, the effect of noise and music in the operating room is complex and not widely investigated, bringing into question any sweeping pessimism of the inclusion of auditory display for image guidance. Unfortunately, clinicians may associate novel auditory display methods with just another alarm. The perception that auditory displays equal alarms could be due to the lack of clinicians’ experience with beneficial auditory displays, but also to approaches that amount to little more than simple alarms applied to an image guidance task.

Parseihan et al. [115] note a major problem of designing auditory display for guidance tasks: the aesthetics of resulting sound design. The authors cite the discomfort caused by using auditory display designs; they can become fatiguing to use or do not match the listening tastes of the intended user. Indeed, the relationship between the intended urgency of a situation and the urgency perceived when using an auditory display is an important issue that should be considered during design [44]. Common auditory warnings in the operating room have been found to be inappropriate, conveying an unintended level of urgency [108]. Thus, clinicians’ dissatisfaction with appropriately urgent alarms may be one reason that investigations into potentially useful auditory display for image guidance never properly develop.

Many of the reviewed approaches are indeed simple in nature, and most articles do not cite psychoacoustically or psychologically driven motivations for sound design decisions, prompting the assumption that most investigations tend to lack interdisciplinary collaboration between researchers of image guidance systems and researchers in field of auditory display.

Bringing auditory display for image guidance into the operating room to provide usable and flexible support for clinicians demands fundamental changes. Enhanced cooperation with sound designers and experts in auditory display to produce more aesthetically appropriate auditory displays will encourage contextual inquiry to help limit implementation in cases when it is disadvantageous, and help discover new applications which might benefit from auditory display.

Increasing sound design complexity so that auditory displays sound more like instruments and less like alarms could increase acceptance [67] and help differentiate the perception of displays from pure alarms that exist in the already noisy environment. Sound designs that are customizable
based on clinicians’ desires are also an interesting option to increase acceptance [67, 115]. Future development should carefully implement toggling so that sound output is only produced when absolutely necessary, a suggestion also offered by Dixon et al. [39]. This will further reduce unnecessary sound output and its related annoyance.

More thorough evaluations of developed methods will help discern exactly which aspects of auditory display are most useful in the operating room and which are superfluous or better supported by other means. In addition to comparing the effects of auditory display to other intraoperative modalities such as augmented reality [61], virtual reality, and conventional navigation, evaluations should include comparisons of multiple auditory display methods, which none of the reviewed literature provided. Investigations into comparing multiple methods for auditory display for 1D guidance tasks suggest expanding such evaluations to 2D and 3D tasks [114]. Such an approach could be taken within the context of multidimensional tracked instrument placement, for instance, for needles [6, 7], aspirators, or continuum robots [26].

2.5 Conclusion

This review of the literature on the use of auditory display for image guidance shows that, despite apparent benefits of augmenting or replacing certain aspects of image-guided interventions with sound information, investigations have been sparse. Positive results include increased risk structure awareness, placement accuracy, and general subjective satisfaction with auditory display, although investigators warn of aspects of annoyance and additional noise in busy operating rooms. There is a need for intensified development and comprehensive evaluation of novel auditory displays that reach beyond simple alerts and alarms to provide clinicians the optimal tool when needed during image-guided interventions.
Chapter 3

Medical Needle Placement


About this chapter

During medical needle placement using image-guided navigation systems, the clinician must concentrate on a screen. To reduce the clinician’s visual reliance on the screen, this work proposes an auditory feedback method as a stand-alone method or to support visual feedback for placing the navigated medical instrument, in this case a needle. This chapter introduces an auditory synthesis model using pitch comparison and stereo panning parameter mapping to augment or replace visual feedback for navigated needle placement. In contrast to existing approaches which augment but still require a visual display, this method allows view-free needle placement. Using combined audiovisual display, participants show similar task completion times and report similar subjective workload and accuracy while viewing the screen less compared to using the conventional visual method.
3.1 Introduction

Image-guided needle navigation is a growing field in which an operator inserts an applicator into a patient’s body to target a tumor or perform a biopsy. For image-guided needle insertion, this can be accomplished using a 2D or 3D visualization on a screen showing the position of a tracked needle in relation to the patient’s anatomy. In some cases, a screen is either unavailable, or the clinician must assume an uncomfortable body position to view a screen [48]. Auditory display may aid the clinician during needle placement using sound cues in addition or instead of the standard visual display. By doing so, the operator receives immediate guidance information to help focus attention on the situs or ameliorate situations in which a screen is unavailable or uncomfortable to view. In contrast to other attempts which solely augment visual feedback, we have developed an auditory display that reaches beyond augmentation and provides auditory feedback that is also capable of screen-free guidance. Our system supports placing the tip and handle of a navigated needle during guidance by means of a parameter-mapping auditory display that employs pitch comparison and stereo placement.

To assess the general applicability of this auditory feedback method, we performed a first evaluation with 12 participants in a laboratory study. This study with novice participants was performed to provide insight into the potential of auditory display and to provide a basis for refined follow-up studies with clinical end users. We compare auditory, visual, and combined feedback for placing a needle on a gelatin phantom under standardized conditions. We investigate the differences with regards to placement accuracy, task completion time, subjective workload, and viewing behavior. This work investigates in particular to what extent auditory feedback differs from both other types of feedback with visual information (combined and visual). Moreover, we investigate how the combined feedback may improve performance and reduce screen viewing time compared to visual feedback.

Because of the challenge of providing complex spatial information with audio and its novelty for most participants, we hypothesize that auditory-only feedback will exhibit higher task completion time, higher subjective workload, and slightly impair placement accuracy compared to both feedback methods providing visual information. Based on results of previous studies [12, 67], for combined feedback, we expect a prolongation of a task completion time, higher accuracy, and reduced subjective workload compared to visual feedback. Moreover, for combined feedback, we expect a reduction in screen viewing time compared to visual feedback.
To the authors’ knowledge, this work demonstrates the first investigation of the impact of auditory-only and combined audiovisual feedback on participant performance for navigated needle placement compared to typical visual-only feedback.

3.2 State of the Art

Modern sound synthesis methods can deliver information in realtime as an alternative or addition to visual systems. To harness this, auditory display can be succinctly defined as a display that uses sound to communicate information. Sonification is the “transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation.” [152]. Hermann [72] proposed that sonification is “[a] technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering).”

Parameter mapping is a form of auditory display that links continuous changes in one set of data to continuous changes in audio parameters. In essence, the underlying data are used to ‘play’ a realtime software instrument according to those changes [72]. Because audio has various parameters that may be altered (such as frequency, intensity, and timbre), continuous parameter mapping is also suitable for displaying multivariate data. This technique often makes the listener an active participant in the listening process by browsing the data set using the auditory display or by interactively changing the mappings that relate data to audio. The listener can navigate through a set of data to perform a task. This method is useful for smoothly representing continuous changes in underlying data.

3.2.1 Auditory Display in Image-Guided Therapy

Auditory display has been employed in image-guided medicine, although its primarily use has been to supplement visual displays and not function as a stand-alone means of navigation. These auditory displays warn the clinician when certain structures have been approached [36, 39, 99, 150, 157, 158]. Compared to no navigation or visual navigation, these works demonstrate benefits including less possible complications [36, 99], better safe margin maintenance [36], higher surgery speed, improved subjective assessment of proximity to critical structures [39, 40, 61], lower subjective workload [39, 40], higher resection precision [157], improved risk structure avoidance and surgical orientation [150], and higher subjective resection quality [158].
Hansen et al. [67] present the first auditory display for image-guided medicine to extend beyond simple warning sounds to support liver resection by guiding a surgical aspirator along a planned resection line on the liver surface. Combining auditory and visual displays led to longer task completion times but increased accuracy and significantly reduced time viewing the screen compared to visual feedback. Purely audio feedback for navigated needle placement has been suggested since 1998 [156], although no extensive evaluations of these systems have been provided.

3.3 Methods

3.3.1 Visual Navigation System for Needle Guidance

Different image-guided navigation systems for liver surgery that track surgical tools in relation to the patient and visualize the patient’s image data have been presented by several groups [9, 33, 118]. In this work, we adapt the commercially available CAS-ONE IR (CAScination AG, Bern, Switzerland) stereotactic needle navigation system. Arnolli et al. [6] provide a generic workflow for percutaneous needle placement that consists of planning a path in image space, retrieving skin entry point, estimating insertion angle, inserting the needle, and performing biopsy or ablation. For the developed method, the system supports the steps of placing the needle at the skin entry point and finding the insertion angle of the handle. Depending on the surgical procedure and procedure step, specific visualization alternatives are utilized (e.g., a 3D volume rendering or a multiplanar reconstruction). For the utilized needle placement, a cross-hair plot displays the relative position of a needle tip and the needle’s orientation for a defined trajectory to a target structure, see Figure 3.1.

To find the correct insertion point of that trajectory, the needle tip position, shown as a red circle, is to be brought to the center of the cross-hair. The correct needle angle to the surface is found when the green circle is in the center of the cross-hair as well. The orientation of the coordinate system is aligned to the needle, such that the positive $y$ direction aligns with the part of the tracked needle facing toward the infrared stereo tracking camera and away from the operator, see Figure 3.2.
3.3. Methods

3.3.2 Auditory Navigation System for Needle Guidance

The developed auditory display for needle guidance is based on a parameter-mapping of the visual feedback, described above, and maps changes in needle placement onto auditory parameters that change in realtime, see Figure 3.3. The auditory display system receives both tip and handle position data in two dimensions sent using Open Sound Control messages across a local network [160]. In fundamental terms, for both tip and handle placement, changes in the $y$ axis (up and down with relation to the operator) are mapped to the moving pitch of tone that the user compares to a static pitch of another tone. These are brought together such that the pitches are the same and $y = 0$. Changes in the $x$ axis are mapped to the stereo position of the tones, such that the tones are brought to the center so that $x = 0$. To minimize possible distraction, synthesized tones for the auditory display method were developed which would not interfere with typical existing operating room sounds, such as ECG or pulse readings from anesthesia monitoring equipment.

These correspond to the visualized tip placement and handle orientation circles shown in the cross-hair of Fig. 3.1. In pilot studies during the iterative design phase of this auditory display method, participants had difficulty perceiving auditory navigation for both tip and handle simultaneously. To simplify cognition, the placement sequence was divided into separate tip and handle placement, so that after correct tip placement, the auditory display is switched to feedback for handle alignment. The $x$ and $y$ unitless values with range and domain of $-25$ to $25$ corresponding to the width and height

![Figure 3.1: On-screen cross-hair visualization, in which the tip and handle of the needle are shown as red and green circles, respectively](image-url)
of the given visualization, are mapped to auditory synthesis parameters including pitch and stereo position. The PureData audio programming environment [120] is used for realtime synthesis. Movement along the $y$-axis is mapped to the pitch of a triangle wave oscillator, which the user compares to a reference oscillator. These pitches of the moving oscillator range from MIDI note numbers\(^1\) 48 to 72 quantized to equal temperament semitones, and the reference pitch has a constant frequency of 261 Hz (MIDI note 60). Thus, for $y$, the frequency $f$ of the moving tone is given as:

\[
    f = 2^{(2y-9)/12} \cdot 440 \text{ Hz}
\]

\(^1\)For a complete table of MIDI note numbers to frequencies, see [159].
3.3. Methods

Following discussions with partner radiologists, playback with head-phones was found to be beneficial. This allows the additional employment of stereo-based parameter mapping, which was found in preliminary experiments to be more effective in conveying right-left navigation than a single monaural channel. For both moving and reference oscillators, phase-offset amplitude modulation at a frequency of 1.8 Hz allows the operator to hear both oscillators alternately to compare pitch. Movement along the x-axis is mapped to the linear stereo position of the oscillators. For \( x \), the gain factors for left and right output channels, \( g_l \) and \( g_r \), are:

\[
\begin{align*}
g_l &= \max \left( \min(x, 1) + 1, 0 \right) \\
g_r &= \min \left( 1 - \frac{\min(x, 1) + 1}{2}, 1 \right)
\end{align*}
\]

Thus, when the tip or handle is located left of the planned position, sound output occurs in the right ear, and when located right of planned position, output occurs in the left ear. Using this mapping metaphor, the user moves a ‘listening object,’ in this case the tip or handle of the needle, towards the ‘sounding object,’ in this case the target position. This frame of reference

![Diagram](image_url)
for stereo mapping means that the user navigates towards the target. When distance of the needle tip or handle to the origin is less than 1, a confirmation tone of a C-major chord plays in addition to the moving and reference triangle wave oscillators.

3.3.3 Experimental Design

We conducted a user study to evaluate auditory (A), visual (V), and combined (C) feedback methods with respect to accuracy, task completion time, time spent viewing screen, and subjective workload during a typical needle placement task. Twelve participants with an average age of 26.25 years took part in the evaluation (five female, seven male, all right-handed). Six participants were medical students in advanced semesters, one was a resident physician, and five were students or employed in other fields. Participants had limited to no navigated needle placement experience.

The task consisted of placing and orienting a navigated needle on the surface of a gelatin phantom (20.0 cm length × 15.0 cm width × 9.6 cm height), see Figure 3.4. This task was divided into two subtasks: placing the tip of the needle at the correct insertion point, and orienting the handle to the correct angle, corresponding to a typical needle insertion work flow [6]. For each placement, the surface of the gelatin phantom was covered in aluminum foil to obscure previous markings. The navigation screen was placed
to the left of a table with the phantom, located approximately two meters from the participant’s standing position, see Figure 3.2. Over-the-ear headphones (Sony MDR-7506) were provided, and participants chose a listening volume that they felt was comfortable.

A within-subject design was used, in which each participants performed the task 9 times, three times under each condition (audio, visual, and combined). Three manually predefined trajectories with varying surface insertion points and handle alignments were used for all participants. The sequence of the three conditions was balanced across participants. After greeting and instruction, participants completed placements for three conditions (audio, visual, combined), consisting of instruction on the method, one unrecorded training placement and two recorded test placements. For each placement, participants placed the tip of the needle, then the handle, declaring ‘start’ and ‘stop’ when beginning and ending each tip and handle placement.

3.3.4 Dependent measures

Four dependent measures were used to assess participants’ performance: First, accuracy of the placement of the needle tip (surface position) and handle (angle), defined as the 2-dimensional Euclidean distance of actual needle position to planned position. Second, task completion time was determined for both subtasks, measured as the time in seconds between declared ‘start’ and ‘stop’ commands given by participants. Third, subjective workload was assessed for each feedback with the modified NASA-Task Load Index, the so-called Raw TLX (RTLX) for which the ratings of different dimensions are averaged to gain an overall workload value [69]. The NASA-TLX assessment tool is a simple method to estimate subjective workload using a multi-dimensional approach that provides information about perceived task demands and subjective reactions to them, although drawbacks to this method include memory effects, response bias, and correlation with task performance. Finally, the fraction of the task completion time spent viewing the screen and phantom was calculated. Participants’ faces were recorded with video, which was manually analyzed by two researchers, whose results were averaged. The correlation between the values of two researchers was > 0.91.
3.3.5 Statistical Analysis

For accuracy and task completion time, the results of both placements were averaged and analyzed by MANOVA (multivariate analysis of variance) for both subtasks as dependent variables and three experimental conditions (audio, visual, and combined feedback) as an independent factor. Subjective workload was analyzed by ANOVA (analysis of variance) with repeated measures with one factor representing the three experimental conditions. We used a common significance level of $\alpha = 0.05$. Furthermore, the data was analyzed by two independent a priori planned contrasts of means. The first contrast addressed effects of audio compared to both feedback types with visual information (averaged across visual and combined). The second contrast compared visual and combined feedback. For accuracy and task completion time, the contrast analyses were performed for each subtask separately and by two-sided t-test for paired samples. Because of this, we corrected a common significance level to $\alpha = 0.025$. Relative viewing time was analyzed by non-parametric Wilcoxon signed-rank test due to a ceiling effect. A common significance value of $\alpha = 0.05$ was used for this test as well.
3.4 Results

The summary of the statistical analysis is presented in Table 3.1.

3.4.1 Accuracy

Effects of three experimental conditions on accuracy in two subtasks are illustrated in Fig. 3.5. Highest accuracy was achieved with the combined feedback, with deviations averaging 1.29 mm (tip) and 1.05 mm (handle), followed by visual feedback, with average deviations of 1.51 mm (tip) and 1.16 mm (handle). Using auditory feedback, the accuracy was slightly lower, average deviations of 1.71 mm (tip) and 1.41 mm (handle), respectively. However, the statistical analyses by MANOVA do not show any significant effects, and the differences to the combined and visual feedback are smaller than the tracking accuracy of the provided system. The following contrast analyses also do not show any significant effects.

3.4.2 Task Completion Time

A significant difference was found with respect to the time needed to complete task, see Figure 3.6. Participants needed significantly less time to place a needle using visual (7.74 s tip, 5.29 s handle) and combined feedback (9.25 s tip, 5.39 s handle) compared to auditory feedback (18.05 s tip, 20.77 s handle), which was reflected in the significant results of contrast 1 (auditory vs. visual and combined) for both subtasks. However, between visual and combined conditions (contrast 2), no difference was found.

3.4.3 Subjective Workload

As seen in Fig. 3.7, a similar pattern was found for subjective workload. Participants reported relatively high subjective workload for audio feedback compared to visual and combined feedback. This pattern could be observed on all dimensions of NASA-TLX, with average values across of individual dimensions of 60.5 for auditory feedback, 34.0 for visual feedback, and 36.0 for combined feedback. This overall result was reflected in the significant effect in the one-way ANOVA which compares these three conditions and the significant contrast 1 comparing audio with visual and combined feedback. No differences appeared between visual and combined feedback (contrast 2).
Chapter 3. Medical Needle Placement

**Figure 3.5:** Accuracy per task in millimeters to planned tip and handle position.

**Figure 3.6:** Time per task in seconds for tip and handle placement.
3.5 Discussion

This work demonstrates the first investigation of the impact of auditory-only and combined audiovisual feedback on navigated needle placement accuracy, task completion time, and subjective workload compared to a typical visual-only feedback method. A parameter-mapping auditory synthesis model was developed to use sound to encode the tip and handle placement of a needle on a gelatin phantom for a medical needle placement task. We investigated whether augmenting existing visual feedback with auditory display offers benefits in increasing accuracy, reducing task completion time, reducing subjective workload, or reducing screen viewing time.

![Subjective Workload as NASA-TLX Scores](image)

**Figure 3.7:** Subjective workload as NASA-TLX scores from 0 to 100

3.4.4 Viewing Duration

Participants viewed to the navigation screen significantly less with the combined feedback, during which 82% of the total time was spent viewing the screen. For the visual-only condition, 99% of the total time was spent viewing the screen.
Following our hypothesis and previous work in path-following in image-guided navigation using a combined auditory display [67], the combined feedback significantly reduced the time viewing the navigation screen. For accuracy of tip and handle placement, no significant differences between three feedback methods emerged. The audio-only feedback was comparable to that of visual or combined feedback and was within an acceptable error of under 2 mm. In contrast to Hansen et al. [67], who reported increased accuracy for combined feedback, the current study did not reach a level of significance for accuracy, possibly due to the increased complexity of the task (2-dimensional as opposed to 1-dimension navigation) and, accordingly, increased complexity of the auditory display.

Task completion time and subjective workload provided a different picture: as hypothesized, ‘blind’ auditory-only feedback led to higher task completion times and increased subjective workload compared to both methods which incorporated the visual display (visual only and combined audiovisual feedback). These deficiencies can be traced to the fact that complex static information can often be processed better visually than with audio. Sounds in nature are linked to status change and motion, whereas visualization can better portray static objects and can be perceived at greater distances. Thus, visual information may give a more complete overview of an environment. In addition, the placing the needle using visual feedback mimicked the typical task of following an object such as a mouse cursor on a screen, whereas using solely auditory display for such a two-dimensional placement task was a new challenge for all participants. Needle placement is a highly complicated task which requires assessing position and orientation of two elements of the object in three-dimensional space and witnessing their change over time. Transmitting this information with audio understandably demands higher temporal and cognitive effort, which is reflected in the results of this study. Advantage with respect to accuracy for combined feedback may likely become present under conditions in which the clinician assumes a body position that would make viewing a screen uncomfortable or impossible, such as guidance of a needle inside an MRI scanner.

Improved auditory feedback mechanisms and increased participant familiarity and training with auditory display may decrease task completion time and further increase accuracy. To our knowledge, the described method presents the first evaluation of auditory-only display for image-guided medical instrument placement. The novelty of the presented method invites a completely new direction for research in auditory display for navigated interventions. Existing approaches all demand that the clinician devote a large amount of time viewing a screen, providing audio mainly as a warning system to inform the clinician of upcoming risk areas. Our method paves a new path by attempting to provide active navigation by audio.
the results of this first study were promising, they show that future designs should exhibit decreased subjective workload. Optimized auditory display methods could be then applied to other applications in medical navigation and reduce screen viewing time substantially. Auditory display for navigation could also encode distance-based risk maps [68] to provide the clinician with a comprehensive solution for both reaching the target and avoiding risk structures. In addition, auditory display for image-guided instrument placement could include an encoding of both navigation information and the uncertainty present therein during targeting tasks [19]. By mapping estimated uncertainty onto the amount of modulation to change certain auditory parameters, such as those described above, a clinician could be informed of the reliability of underlying navigation data.

Apart from auditory displays, several methods have been described to encode information for navigated interventions using visual augmented reality (AR) [86], e.g., using video see-through displays, optical see-through displays, or projectors to augment the operation with navigation information directly on the patient. However, AR may be accompanied by drawbacks including attention tunneling [40, 102] which may jeopardize patient safety. Future research could investigate how such advanced medical augmented reality visualization techniques could be beneficially combined with auditory feedback, for example, to improve intraoperative assessment of depth and distances, especially for cases in which such novel video displays could compromise patient safety.

### 3.6 Conclusion

This study evaluated auditory, visual, and combined feedback methods in comparable standardized conditions to determine the differences in their effects on performance in a typical image-guided medical needle placement task. By adding auditory display to existing visual display, time spent viewing the screen was significantly decreased while maintaining accuracy and task completion time. After refining this auditory display method, an evaluation of the concept will focus both clinical trials as well as laboratory studies in which the navigation system display may be difficult to see due to operator body position.
We have shown that auditory feedback can be employed in laboratory conditions as an augmentation to visual feedback or employed as a singular feedback for situations in which a screen is unavailable, uncomfortable to view, or a high value is placed on being able to view the operating situs. Currently, however, the results suggest that auditory displays that reach the performance of visual-only displays are still to be developed, emphasizing the current need for visualization during interventions. Future research should place intensified consideration of auditory as a navigation feedback in a field where visual-only guidance is dominant.
Chapter 4

Telerobotic Transnasal Surgery


About this chapter

Tubular continuum robots can follow complex curvilinear paths to reach restricted areas within the body. Using teleoperation, these robots can help minimize incisions and reduce trauma. However, drawbacks include the lack of haptic feedback and a limited view of the situs, often due to camera occlusion. This work presents novel auditory display to enhance interaction with such continuum robots to increase accuracy and path-following efficiency and reduce cognitive workload. A test environment simulates a typical transnasal intervention through the sphenoidal sinus using a continuum robot. Distance information is mapped to changes in a real-time audio synthesizer using sung voice to provide navigation cues. User studies with novice participants and clinicians were performed to evaluate the effects of auditory display on accuracy, task time, path following efficiency, subjective workload, and usability. When using auditory display, participants exhibit significant increases in accuracy, efficiency, and task time compared to visual-only display. Auditory display reduced subjective workload and raised usefulness and satisfaction ratings. The addition of auditory display for augmenting interaction with a teleoperated continuum robot has shown to benefit performance as well as usability. The method could benefit other scenarios in navigated surgery to increase accuracy and reduce workload.
4.1 Introduction

Tubular continuum robots are characterized by their non-linear shape and small diameter (< 2.5 mm). For medical interventions, this enables manipulation of the robots via non-linear complex paths through natural orifices to restricted areas in the situs. These robots consist of biocompatible, flexible, pre-curved tubes nested into one other, which are specifically chosen to meet the application requirements and patient anatomy. These form the so-called backbone of the manipulator. For a thorough review of the current state of research of tubular continuum robots, see [55]. These flexible robots are meant to be used with control via teleoperation in most minimally invasive applications by the surgeon, who manipulates a master input device [26, 96]. Using this input method, the robot manipulator (slave) follows the commands of the master, who interacts with the environment through the robot.

In general, teleoperated surgical robots can enhance, for instance, minimization of incisions and trauma as well as reduce the physical load on the surgeon to allow natural hand-eye coordination, motion scaling, and high-end filtering of natural tremors [144]. Thanks to these characteristics, tubular continuum robots have been suggested for many different clinical and medical applications; for a general overview, see [28]. For example, Burgner et al. proposed the use of tubular continuum robots for intracerebral hemorrhage evacuation [27]. Other proposed applications of tubular continuum robots include use in intracardiac [58] and neurosurgical [5] interventions, all of which benefit from the robot’s dexterity and miniaturization to allow minimally invasive surgery. Furthermore, teleoperated tubular continuum robots have been explored in transnasal skull-base surgery scenarios [26]. In this case, a tubular continuum robot is deployed through the transnasal access in order to reach tumors located at the skull base. The curvilinear shape of the robot enhances tumor reachability compared to straight and rigid tools. In this paper, we will, as in other studies, concentrate on transnasal skull base surgery as a use case to evaluate our developed methods.

Despite the advantages of tubular continuum robots, there are still drawbacks with current approaches. In most cases, interventions are performed with an endoscope, causing the surgeon to lose the natural feeling of the forces of interaction with tissue structures that are normally exerted during the procedure. Additional challenges of teleoperation using surgical continuum robots include a limited view of the surgical site, a lack of visual depth information: commonly available 3D endoscopes do not physically fit inside the central cavity of the innermost tube diameter of 0.2 to 2.5 mm. In addition, there is an increased cognitive load on the operator due to the operation of a curvilinear instrument. Furthermore, when the view of the endoscope is obscured, for instance by blood, navigation becomes more difficult.
To overcome some of these challenges, this work proposes employing results from the growing field of auditory display to provide guidance information through a hitherto relatively free sense. A so-called auditory display maps changes in an underlying set of data. A well-known example of an auditory display used to transmit motion changes can be heard in common parking assistance systems: the distance of the front or rear of an automobile is mapped to the inter-onset time between tones, culminating in a steady tone once the automobile reaches a defined threshold to the other parked automobiles or obstacles. In a similar fashion, auditory displays for medical robot guidance can be used to hear changes in navigated instruments for guidance. In this case, changes in motion of a tracked continuum robot are mapped to changes in parameters of a real-time sound synthesizer.

The primary motivations for developing auditory display for medical robot guidance include increasing clinician’s awareness of surrounding structures, replacing the diminished sense of touch, and assisting clinicians in correctly interpreting and more accurately following navigation paths. Although examples of auditory display in medical navigation are scarce, methods have been evaluated for radiofrequency ablation [18], biopsy needle placement [22], temporal bone drilling [36], skull base surgery [39], soft tissue resection [67, 157], and telerobotic surgery [90]. Previous evaluations of auditory displays for medical navigation have shown benefits including enhanced recognition of the distance to structures or targets, improved placement accuracy, reduced cognitive workload, and reduced rates of clinical complication. Drawbacks reported in the literature include increased task time and increased non-target tissue removed during resection. For a comprehensive review of auditory display in image-guided interventions, see Black et al. [16].

Previous work on teleoperation algorithms to assist continuum and surgical robots have focused on either automating several subtasks of teleoperation, such as obstacle avoidance, or providing the master with additional feedback for further support. For example, Leibrandt et al. [96] applied virtual fixtures to concentric-tube continuum robots. They proposed teleoperation algorithms that guide the user along stable and collision-free paths using haptic feedback. Kreuer et al. [56] implemented and evaluated a similar approach using a simulated transnasal endoscopic surgical intervention. Results have shown that in comparison with purely visual feedback, haptic feedback generated using virtual fixtures enables more intuitive teleoperation while also lowering the completion time of the presented tasks. Apart from providing additional feedback, some algorithms for teleoperation assistance also directly affect with the motion generation of the continuum robot. For example, Torres et al. [145] proposed a system for concentric-tube continuum robots that allows teleoperation of the robot’s tip while helping avoid
collisions along the robot’s entire backbone. Furthermore, Gras et al. [59] have shown that an adaptive motion scaling during teleoperation of surgical robots can lead to more intuitive teleoperation due to less reliance on the ‘clutch in’-mechanism.

This work describes a method to harness the benefits of auditory display to augment the task of virtual tubular continuum robot placement using a haptic input device. Tasks include 1) locating waypoint spheres inside a transnasal access passage and 2) avoiding the passage’s walls. These two tasks are evaluated in a user study to compare the quantitative and qualitative benefits of using such a system with and without auditory display. In many clinical scenarios, image guidance navigation assistance for a tracked instrument can be given using three parameters: azimuth (left and right), elevation (up or down), and depth. This can be applied, for instance, to needle placement [48], image-guided laparoscopy [79, 161], or transnasal telerobotics [27], where an instrument must be navigated to remain on the origin of a plane orthogonal to the line to a planned target. Thus, similar to navigation inside a tunnel, the instrument must remain along a trajectory until the target in depth has been reached. To the authors’ knowledge, no auditory display has not been explored yet for applications involving medical tubular continuum robot navigation. Thus, this work goes beyond the state of the art to increase the understanding of the role audio can play in hybrid-display clinical navigation systems, especially those in which tunnel-like scenarios play a large role. We hypothesize that the addition of auditory display will increase accuracy, efficiency, and usability of continuum robot navigation as well as reduce cognitive workload for a task involving continuum robot navigation.

4.2 Methods

In this work, changes in motion of an input device controlled by the operator are transmitted to the motion of a virtual tubular continuum robot composed of three tubes. The changes in motion of the robot are transmitted to the auditory display synthesizer, which produces sound output for the user to hear.
4.2. Methods

4.2.1 Robot System

Each tube of the tubular continuum robot has two degrees of freedom. Typically, these robots are composed of three tubes, such that a configuration \( q \) consists of translational \( [\beta_1, \beta_2, \beta_3] \) and rotational parameters \( [\alpha_1, \alpha_2, \alpha_3] \). The quasi-statics mechanics-based model of the robot is based on [126]. The model determines the solution to the forward kinematics for a given configuration \( q \) as a continuous 3D curve describing the shape of the robot in terms of its centerline. In addition, the model is used to determine the manipulator’s Jacobian \( J \). The kinematics model is based on the specific case of Kirchhoff of the Cosserat rod theory and may involve forces acting on the robot.

The input device (Geomagic Touch, 3D Systems [1]) is an impedance-controlled input device consisting of 6 degrees of freedom. It is controlled by a hand-held pen stylus for which the position and orientation in Cartesian space are detected at the pen tip. Velocity is calculated from the generated change in position of the pen tip to the desired end effector. Due to the different workspaces of the input device and of the robot, a so-called ‘clutch-in’ mechanism is applied by pressing a button on the stylus to move the end effector of the robot. By releasing this button, the stylus can be repositioned without end effector movement.

The motion of the input device commanded by the operator is expressed as a desired end effector velocity \( \dot{x}_{\text{des}} \) (position and angular velocity) dependent on the current continuum robot pose \( x_{\text{robot}} \), which is obtained using the kinematic model description described in [126]. Then, the pseudoinverse of the manipulator Jacobian matrix \( J^+ \) is used to calculate the corresponding...
Chapter 4. Telerobotic Transnasal Surgery

joint velocities $\dot{q}$ [127]. These are summed up to the current joint values $q$ of the robot in order to move towards the desired position $\mathbf{x}_{\text{des}}$ and to update the end-effector position prediction based on the kinematic model. This process is shown in Fig. 4.1.

4.2.2 Simulation Environment and Task Description

The simulation environment includes an anatomical model of the transnasal access which was manually segmented from a computed-tomography (CT) dataset of an adult skull base surgery patient. In the implemented scenario, the robot’s end effector should be teleoperated through the transnasal access to the sphenoidal sinus. This allows the manipulator to reach the patient’s skull base, which is a common location of pituitary tumors. For the implemented teleoperation environment, we decided for a motion scaling factor of 0.1, resulting in a robot end-effector movement ten times smaller than the movement at the input device (i.e., moving the input device by 1 cm leads to a end-effector movement of 1 mm). A visualization of both the robot and the anatomical model can be seen in Figure 4.2. The camera view within this simulation mimics the camera view of an endoscope which is typically guided alongside the robot through the transnasal access.

During robot teleoperation, we consider two different use cases. For the first, a path from the entrance of the transnasal access to the sphenoidal sinus consisting of several discrete waypoints (in Figure 4.2 shown as blue spheres). This path can be precomputed by a motion planning algorithm and serves as guidance for the teleoperating user, who steers the robot’s end effector to each waypoint successively until the final waypoint has been reached. The second use case does not feature waypoints. For this task, the user navigates the robot’s end effector towards the end of the transnasal corridor while avoiding collision with the walls of the anatomical model.

For both use cases, we propose that auditory display could assist the user in navigation in addition to the visual feedback of the simulation environment. For the use case of navigating waypoint spheres, the distance in the $x$- and $y$- directions between the robot’s end effector and the next waypoint sphere with respect to the robot’s base frame is calculated at every step of the teleoperation algorithm (see Fig. 4.3, left). In addition, the depth between the next waypoint and the robot’s end effector is calculated. For the use case of collision avoidance, the discrete data points of the virtual anatomic model are stored in an octree structure.
4.2. Methods

This search tree structure can query collisions and distance checks based on the Euclidean distance between the queried point and the data points stored within this tree [105]. Using this octree, a radius search around the robot’s end effector is performed at every step to find nearby anatomical data points. The \( x \)- and \( y \)- distances of the anatomy with respect to the robot’s end effector are calculated by averaging the distances of all found data points within a specified radius (see Fig. 4.3, right). For this implementation, we chose 3.5 mm as radius. A distance between the robot and the anatomy greater than this specified radius can be considered as safe in the context of this application.

4.2.3 Auditory Display for Robot Navigation Assistance

For each of the two calculated metrics, the \( x \)- and \( y \)- distances to the next waypoint sphere (in mm) or the \( x \)- and \( y \)-distance (in mm) to the closest wall, an auditory display was developed to guide the user during both teleoperation use cases. The general approach for developing an auditory display for the continuum robot is based on a mapping the three degrees of freedom in manipulator end-effector position to sound: azimuth from -10 mm (left)
Chapter 4. Telerobotic Transnasal Surgery

Figure 4.3: Left: Calculation of the x and y distances between the next blue waypoint spheres of the path and the robot’s end effector. Right: Radius search around the robot’s end-effector to find nearby anatomical data points (simplified as dots). Dots inside a specific radius $r$ (marked in red) are averaged to calculate the $x$ and $y$ distances to the anatomical model.

to +10 mm (right), elevation from -10 mm (below) to +10 mm (above), and depth (100 mm away from target to 0 mm at target). Throughout this work, no feedback based on the orientation of the manipulator is given as the task only requires position control.

The auditory display method was developed using the PureData [120] visual music and sound programming environment. The development of the auditory display followed an iterative design process that included multiple rounds of hands-on preliminary evaluation with novice and expert users. In these preliminary iterations, some participants were irritated purely synthesized ‘electronic sounds’, such as those similar to methods developed in previous studies (e.g. [18]), and thus, an alternative method was developed to reduce unnecessary additional cognitive load resulting from annoyance. The subsequently developed and evaluated auditory display method described here employs a set of recorded samples of vocal singing to provide navigation cues. A ‘vocal syllable choir’ synthesizer responds to the end-effector position. Deviations from planned elevation, azimuth, and depth
are mapped to parameters of the auditory display synthesizer. The synthesizer harnesses pitch, syllable type, and voice gender (for elevation mapping), stereo position (for azimuth mapping) and reverberation amount and inter-onset interval (for depth mapping), see Fig. 4.4 for a simplified representation of employed mappings.

Deviations in elevation are mapped to the pitch, syllable type, and gender of the sung syllables for redundant cognition. Categories of syllables were first created, for which sung syllable samples were recorded. This harnesses the powerful ability of humans to recognize syllables. The categories of syllables included [ɔː:] for deviations above the x-axis (German: oben), [uː:] (German: unten) for deviations below the x-axis trajectory, and hummed [m] when elevation was correct within a defined margin to create a steady confirmation tone. Pitches for deviations of the end effector below the x-axis range from E2 (82.41 Hz) at -10 mm to C3 (130.81 Hz) at -1.1 mm. Pitches for deviations above the x-axis range from C4 (261.63 Hz) at 1.1 mm to B4 (493.88 Hz) at 10 mm. For deviations within the defined margin (‘correct elevation’) deviations between -1 mm and 1 mm, a triad with hummed pitches E3-G3-A3 (164.81/196.00/440.00 Hz) is played.

Deviations in azimuth from the center are mapped to the stereo position of the output. In pilot studies as well as previous work [18], this was found to be an effective way of transmitting current azimuth compared to monaural presentation. Thus, when the tip of the robot is located to the left of the center, sound output is heard in the left ear. When the tip is located to the right of the center, sound output is heard in the right ear. Between -1 mm and 1 mm, output is heard with a linear mapping between left and right ears. Outside of this distance, output is heard completely in the left or right ear to aid discrimination and enhance fine azimuth adjustment near the center.
For the waypoint navigation task, deviations in depth are mapped to the inter-onset interval of the sung samples as well as the amount of reverberation mix. At the furthest distance (15 mm) to the next sphere, the inter-onset interval is 450 ms, which increases to 250 ms when at the sphere. Because each sung syllable sample is approximately 1 second in duration, samples are overlapped, creating a choir effect with up to 8 samples played back at once. The amount of reverberation mix (using the Freeverb object [136]) is 30 % at the furthest depth (15 mm) and 0 % when at the target sphere. The overall resultant sound is that of a medium-sized choir with the ability to represent changes in azimuth, elevation, and depth.

4.2.4 Experimental Design

We conducted two user studies to evaluate the effect of auditory display as augmentation for navigating a virtual tubular continuum robot through the sphenoidal sinus to the pituitary gland. We completed two studies: the first with a group of students and scientific personnel recruited from the local university, and the second with clinicians that featured a smaller set of tasks. Compared to visual feedback, we hypothesized that using the auditory display will:

H1: increase accuracy during navigation
H2: increase task completion time
H3: increase path following efficiency
H4: reduce overall task load
H5: increase overall usability

The two conditions differed only in the presentation of navigational information:

V: only visual display on the screen
A: additional auditory display

We employed a within-subject design and controlled for potential learning effects by using a Latin-square scheme for the order of conditions. Based on the two aforementioned scenarios of teleoperational sphenoidal sinus surgery, the two tasks included:

N: Navigation of consecutive waypoint spheres in sphenoidal sinus
C: Collision avoidance of sphenoidal sinus passage walls

\(^1\) An increase in task completion time has been observed in previous approaches ([16, 18, 67]), and, thanks to the novelty of using auditory display, we also predict such an increase.
This resulted in four task/condition combinations: NV (navigation with visual display), NA (navigation with auditory display), CV (collision avoidance with visual display), and CA (collision avoidance with auditory display).

During task execution, we recorded the position of the robot’s end effector, the position of the nearest anatomy with respect to the robot’s end effector as well as the corresponding Euclidean distance between them. We calculated accuracy of the tip, path efficiency, task completion time. A questionnaire provided insight into perceived workload, system acceptance with regards to usefulness and satisfaction, and participant agreement with confidence, ease of use, time needed, and helpfulness of each task/condition combination. In addition, we asked the participants to rank the overall usability of each of the task/condition combinations.

The waypoint navigation (N) task consisted of guiding the robot along a route delineated by six blue spheres. For this task, participants were instructed to navigate the robot towards and hit the center of each sphere. Accuracy was measured as the Euclidean distance of the end effector to the center of the blue sphere when passing it. Task completion time was determined by the elapsed time between the moments the participant crossed one blue sphere to the consecutive sphere. An efficiency score based on ISO/IEC 9126-4 [76] was calculated based on the accuracy and task completion time as \( \frac{t}{dev^2} \) where \( dev \) is the deviation as the absolute distance from the tip of the robot to the sphere when passed, and \( t \) is task completion time, the elapsed duration needed between two consecutive spheres. This appropriately weights the accuracy higher than the task time, which was generated in discussions with partner clinicians. Alternating spheres were occluded by simulated blood to create an obstacle that might occur in a real operating situation.

The collision avoidance (C) task consisted of guiding the robot through a tunnel without colliding with the walls, which represented fragile structures in the sphenoidal sinus. Similarly to the navigation task, simulated blood was displayed in the 3D environment. Accuracy was determined as the average distance to the walls of the anatomical sphenoidal sinus model. Task completion time was determined by the time needed to proceed through the entire transnasal corridor.
4.2.5 Evaluation Procedure

Two studies were performed, one with novice participants (Study 1) and one with clinicians (Study 2). The studies took place in an enclosed laboratory. The test workstation consisted of a computer with a 24 inch monitor and a Geomagic Touch input device [1]. Over-ear studio headphones were provided for each participant. Participants were visually separated from each other and were seated during the entire evaluation procedure. Participants were permitted to adjust chair and input device position and rest their arm on the testing table.

Participants received a 5-minute tutorial video and a 1-page explanation 2 days before the study to familiarize themselves with the study tasks and the auditory display design. After giving informed consent, the participants were assigned according to the Latin-square counterbalance scheme to one of the tasks and conditions to start with. The evaluation was conducted using the aforementioned tasks. Before each task/condition combination, the participants received task-specific training and performed a single training run. Participants were instructed to complete all tasks as quickly and accurately as possible, finding their own trade-off between time and accuracy.

After completing each of the four task/condition combinations, the participants completed a questionnaire about their experience. The questionnaire included NASA-TLX [31], which measures the cognitive workload experienced by the participants while completing the task. We employed the Raw TLX scale, as the weighted scale involves a greater amount of time to complete and has not been conclusively shown to produce extra benefits [69].
4.3. Results

In addition, the commonly employed van der Laan technology acceptance scale [94] was given to evaluate the usefulness of and satisfaction with using the teleoperation system with each of the task/condition combinations. This scale includes 9 pairs of adjectives including “undesirable/desirable” or “nice/annoying” to generate composite ratings for usefulness and satisfaction. Finally, four questions were asked concerning participants’ agreement with confidence in executing the task, ease of use, satisfaction with the time needed to complete the tasks, and helpfulness of the feedback method. At the conclusion of the study, the participants completed a questionnaire concerning demographics (age, handedness, visual and hearing impairments, video gaming frequency) as well as a single ranking of the participant’s preference for each of the task/condition combinations.

4.2.6 Statistical Analysis

The data was analyzed using R [121]. For all data gathered from Study 1, i.e., task completion time, accuracy, efficiency, NASA TLX score, and van der Laan scores, the assumption of normal distribution was rejected by the Jarque-Bera test and therefore all were analyzed with the Wilcoxon signed-rank test. The level of significance, i.e., the statistical difference of the means, was indicated by $p < .05$. For Study 2 we did not perform a inferential statistical analysis, as the sample size is too small to gain meaningful outcomes in terms of significance. Therefore, we present averages and box-and-whisker plots as a descriptive approach to show the magnitude of the effects.

4.3 Results

Study 1 was completed by 16 novice participants. The participants had an average age of 28.1 (ranging from 23 to 31). The participants were students recruited from Leibniz University in Hanover, Germany, and all had a background in scientific fields of study. Eight wore corrective lenses. None of the participants stated any limitations regarding their hearing abilities. One of the participants was left-handed but performed the evaluation with the right hand; the remaining were right handed. Asked about their gaming frequency 31 % (5) of the participants stated that they game about once per month, 24 % (4) weekly, 19 % (3) a few times a year, and 25 % (4) less than once a year.

Study 2 was completed by 5 participants. The participants had an average age of 32.2 (ranging from 19 to 51), with 1 female and 4 males. The participants were recruited from the International Neuroscience Institute in Hanover, Germany. Four were neurosurgeons and 1 was a medical student.
All wore corrective lenses. None of the participants stated any limitations regarding their hearing abilities. All participants were right-handed. Asked about their gaming frequency, 20% (1) of the participants stated that they game a few times a week, while 80% (4) play less than once a year.

### 4.3.1 Accuracy

**Study 1** For the waypoint navigation task, the average absolute distance (error) in the x-y plane of the end effector to the passed sphere was measured in millimeters. Participants performed with significantly less error using auditory display (NA) compared to visual feedback (NV), see Figure 4.6a. For visible spheres, the average error was 1.08 mm using visual feedback and 0.89 mm using auditory display. For hidden spheres, the average error was 3.43 mm using visual feedback and 1.49 mm using auditory display. For the collision avoidance task, the average distance to the anatomical model was similar for both conditions: 3.00 mm for visual feedback and 2.97 mm for auditory display, although this did not reach a level of significance.

**Study 2** Participants in Study 2 performed similarly to those in Study 1, see Figure 4.6b. For visible spheres, the average error was 0.92 mm using visual feedback and 1.04 mm using auditory display. For hidden spheres, the average error using visual feedback 5.66 mm using visual feedback and 2.11 mm using auditory display.

### 4.3.2 Task Completion Time

**Study 1** For the navigation task, the average time to hit each sphere was measured in seconds, with participants needing significantly longer using auditory display (NA) compared to visual feedback (NV), see Figure 4.6c. For visible spheres, the average task completion time per sphere was 7.60 s using visual feedback and 11.21 s using auditory display. For hidden spheres, the average completion time per sphere was 8.03 s using visual feedback and 15.57 s using auditory display. For the collision avoidance task, the task completion time was longer using auditory display (CA), with an average of 79.94 s than visual feedback (CV) with an average of 61.08 s, but did not reach a level of significance.
4.3. Results

Figure 4.6: Average error (in mm), task completion time (in s), and efficiency (mm/s²) per sphere over all participants during the waypoint navigation task, where V-V and V-A stand for visible spheres using visual and auditory feedback, and I-V and I-A for invisible spheres using visual and auditory feedback.
Study 2 For the waypoint navigation task (see Figure 4.6d), the average task completion time per visible sphere was 17.39 s using visual feedback and 18.54 s using auditory display. For hidden spheres, the average completion time per sphere was 18.89 s using visual feedback and 28.78 s using auditory display.

4.3.3 Efficiency

Study 1 Participants performed significantly more efficiently using auditory display compared to visual feedback when approaching hidden spheres, see Figure 4.6e. For these spheres, the average efficiency time per sphere in mm/s² was 0.08 using visual feedback and 0.28 using auditory display. For visible spheres, the average efficiency per sphere was 0.72 using visual feedback and 2.20 using auditory display, although a level of significance could not be reached.

Study 2 Study 2 participants performed more efficiently using auditory display compared to visual feedback when approaching hidden spheres, see Figure 4.6f. For visible spheres, the average efficiency per sphere was almost equal using both visual and auditory display.

4.3.4 Subjective Workload

Task load was calculated as the overall task load index according to the NASA Raw TLX questionnaire [31, 69], the results of which are presented in Figures 4.7a and 4.7b. The subjective workload is presented as an overall scores ranging from 0 (best) to 100 (worst). In Study 1, for both navigation and collision tasks, using the auditory display was rated with significantly lower overall subjective workload than using only the visual feedback. In addition, for every dimension of the index and for each task, the auditory display was rated with significantly lower subjective workload. In Study 2, using auditory display was rated with lower overall subjective load than with visual display.

4.3.5 System Acceptance

Figures 4.7c, 4.7e, 4.7d, and 4.7f present the results of the Van der Laan system acceptance questionnaire [94] as both composite usefulness and satisfying scales, with combined scores ranging from $-2$ (fully reject) to $+2$ (fully accept). In Study 1, for both tasks, both the usefulness as well as satisfaction
4.3. Results

Figure 4.7: Questionnaire scores for waypoint task for visual (NV) and auditory display (NA) conditions and collision avoidance for visual (CV) and auditory (CA) display. NASA Raw TLX ranges from 0 (low workload) to 100 (high workload), for van der Laan from -2 (fully reject) to +2 (fully accept).
when using auditory display were rated as significantly higher than when using visual feedback. In Study 2, both the usefulness as well as satisfaction when using auditory display were rated higher than when using visual feedback.

### 4.3.6 Participant Agreement

The results of participant satisfaction with confidence during use, ease of use, time needed to complete the tasks, and helpfulness of the display are shown in Table 4.1, with scores on a 7-point Likert scale [98] ranging from 1 (complete agreement) to 7 (complete disagreement). For Study 1, for both navigation and collision avoidance tasks, using auditory display was rated as providing significantly higher satisfaction in all measures. For Study 2, all participants rated auditory display higher than visual display for the waypoint navigation task.

<table>
<thead>
<tr>
<th>Measure</th>
<th>NV</th>
<th>NA</th>
<th>CV</th>
<th>CA</th>
<th>NV</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>3.88</td>
<td>2.13</td>
<td>2.75</td>
<td>1.94</td>
<td>6.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Ease of use</td>
<td>3.12</td>
<td>1.81</td>
<td>2.56</td>
<td>1.80</td>
<td>5.40</td>
<td>2.00</td>
</tr>
<tr>
<td>Time needed</td>
<td>2.38</td>
<td>2.00</td>
<td>2.38</td>
<td>1.81</td>
<td>3.20</td>
<td>2.20</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>4.50</td>
<td>1.81</td>
<td>4.63</td>
<td>1.75</td>
<td>5.75</td>
<td>1.40</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>3.47</strong></td>
<td><strong>1.94</strong></td>
<td><strong>3.08</strong></td>
<td><strong>1.83</strong></td>
<td><strong>5.09</strong></td>
<td><strong>1.90</strong></td>
</tr>
</tbody>
</table>

**Table 4.1**: Mean agreement questionnaire values for all conditions, from 1 (strongly agree) to 7 (strongly disagree).

### 4.3.7 Overall Ranking

For Study 1, participants were asked at the end of completing all tasks to assign an overall rank of each of the 4 task/condition combinations from least to most preferred, receiving 0 to 3 points. The average rating scores for each combination are shown in Figure 4.8. CA was preferred the most, with an average rank of 2.31, NA with an average of 1.81, CV with an average of 1.13, and NV with an average of 0.75. For Study 2, participants ranked only NV and NA, and all 5 preferred NA.


\[ \text{Figure 4.8: Average overall ranking scores for each task/condition combination for Study 1, from 0 (least preferred) to 3 (most preferred)} \]

\[ \begin{array}{c|c|c|c}
    & NV & NA & CA \\
\hline
    Total score & 0.0 & 1.0 & 3.0 \\
\end{array} \]

4.4 Discussion

For the waypoint navigation task (N), results of the user study confirm all of our hypotheses. In Study 1, accuracy (H1) in hitting waypoint spheres was significantly improved using auditory display (NA) over visual display (NV), showing that the auditory display provides exceptional performance benefit in accuracy not only when the user cannot see the target on the screen, but also during navigation of visible spheres. Interestingly, in Study 1, average accuracy for invisible spheres was even higher using auditory display than for visible spheres using visual display. Accuracy in path following is important because critical structures should be hit as little as possible. The calculation of accuracy by means of a waypoint task helps recreate the medical scenario in which the walls of the transnasal corridor should not be hit.

Our hypothesis for task completion time (H2) was confirmed; the tasks took significantly longer using auditory display than with visual display, which was expected due to similar results in the literature [16, 18, 67] that have shown that performance with the addition of auditory display entails significantly more time. During post-experimental discussions with study participants, we received feedback that this could be due to the novelty of the introduced method: many users are already familiar with visual displays (spending hours each day interacting, for instance, with mouse and cursor on a computer screen) but less so with auditory displays, where the most typical interaction might be car parking aids, which are used, at most, a few seconds per day. Although accuracy results were similar between both Study
Chapter 4. Telerobotic Transnasal Surgery

1 and Study 2, task completion time for Study 2 was higher for all conditions, although relative times between conditions were similar to those in Study 1, see Fig 4.6d. This is most likely due to the fact that clinicians placed more emphasis on accuracy at the expense of longer task time.

Our hypothesis concerning efficiency in path following (H3) was partially confirmed; auditory display provided significant increases in efficiency for invisible spheres and non-significant increases for visible spheres.

For all measures gathered using the questionnaire (NASA TLX Workload, van der Laan system acceptance, and agreement with confidence, ease, time, and helpfulness), auditory display provided significantly improved scores, thus confirming both H4 (reduce overall task load) and H5 (increase usability). Mean values for both overall workload and system acceptance as well as all individual factors showed that the auditory display improved interaction. In addition, using auditory display for waypoint navigation was ranked higher overall than visual display by 12 of the 16 participants. The cohort of physicians in Study 2 exhibited similar performance to those in Study 1, although average task time was higher.

For medical participants, the difference in subjective measures (NASA TLX, van der Laan, and agreement questionnaires) between auditory display and visual feedback were much greater for all scales than participants in Study 1.

For the collision avoidance task (C), performed only in Study 1, significant differences were not found in either the task completion time or accuracy. Thus we are unable to confirm H1 and H2 for this task. These results might be due to the setup environment of the collision avoidance task, as all participants were able to complete the task without any great difficulty. However, using auditory display was shown to provide significant benefits with respect to both overall NASA TLX subjective workload and van der Laan system usability in both usefulness and satisfying scales, confirming H4 and H5. In addition, 13 of the 16 participants in Study 1 ranked the interaction with the auditory display for the collision avoidance task higher than interaction with visual display. These results indicate that despite the lack of significant performance improvements in accuracy or task completion time, the auditory display created a more agreeable means of robot interaction and was preferred by most participants. Due to the lack of significant results from Study 1, we elected not to perform the collision tasks (CA and CV) with clinicians, due to time and resource constraints.

Overall, we confirm the hypothesized benefits of using auditory display for supporting interaction with a continuum robot for the virtual sphenoidal sinus intervention scenario. Although participants only completed a brief training phase, the auditory method has been shown to benefit a task that is
hitherto performed with only visual display. It is has been demonstrated to provide completely ‘blind’ (screen-free) instrument placement, as well as to increase accuracy and efficiency as well as decrease subjective workload and improve satisfaction. The auditory display also helped improve interaction even when significant performance benefits were not exhibited, such as during the collision avoidance task. Although task times were, as hypothesized, longer when using auditory display, we believe that with increased training with and exposure to this method and auditory displays in general, task completion time could be reduced. The differences between the results of Study 1 and Study 2 were pronounced: whereas novice participants in Study 1 obtained predominantly better performance results than those in Study 2, the clinicians of Study 2 reported higher scores for auditory display, suggesting that such a method would garner high levels of acceptance by operating clinicians. With increased training, the performance levels of clinicians could hypothetically also reach those of novice users. In addition, the growing popularity of video gaming has shown positive effects on surgical task completion time and error rate [125]. In our results, average task completion times for participants who play video games more than a few times per month were considerably lower (9.27 s per sphere for gamers vs. 16.40 s for non-gamers), and error rates were slightly lower (1.81 mm per sphere for gamers vs. 1.98 s for non-gamers). Finally, in consultation with our partner clinicians, we have determined that for this type of scenario, the use of open, over-the-ear headphones would not present ergonomic or performance challenges.

4.5 Conclusion

This work presents an auditory display method developed to aid a broad spectrum of image-guided medical navigation approaches and focuses on two specific use cases of transnasal continuum robot navigation. Thus, the results of this work should help determine whether the addition of auditory display should be further investigated to aid medical tracked instrument navigation of a tunnel-like scenario, also found in clinical applications such as laparoscopy, needle placement and skull base surgery. Tunnel-like navigation is a cognitively demanding application for mapping using auditory display; many degrees of freedom must be simultaneously heard to reach the clinical target. We performed an evaluation with novice and medically trained participants to determine the benefits and drawbacks of using auditory display compared to visual display for tunnel-like navigation in the sphenoidal sinus.
The auditory display method has been shown to benefit instrument placement to reach targets within a sphenoidal sinus scenario. User studies with both novice and medical participants showed that the auditory display provided significant improvements in both quantitative performance measures and qualitative usability measures. Although the methods must still undergo further evaluation in a clinical setting, results from our work show that the use of auditory display for continuum robot navigation will certainly play a role in future mixed- and augmented-reality medical navigation systems.

Future work will focus on refining the chosen auditory display method. A hybrid method of synthesized tones and natural voice samples could provide more intuitive directional cues given by the virtual choir as well as more fine-grained navigation given by additional synthesized tones. In addition to determining optimal auditory display mappings, increased focus should be placed on the role of auditory display in the operating room, especially with regards to augmenting existing visual methods. To the authors’ knowledge, no research has been performed in the medical domain that attempts to determine the optimal combination of auditory and visual feedback for hybrid navigation such as the one presented in this work. Supplementing existing visual navigation aids, for example, with auditory depth cues or distance-based risk warnings could enhance interaction with the system. In this work, we describe an auditory display method developed to support this type of navigation. The described auditory display for tunnel-like navigation aid could, however, also be used to guide tasks with fewer degrees of freedom, such as a liver resection line (as employed by [67]) or absolute distance to risk structures (as seen in [36]).

In summary, this work describes an auditory display method which can be applied to a wide range of clinical use cases involving image-guided navigation. We confirm our overall hypotheses that the addition of auditory display increases accuracy, efficiency, task completion time, and usability, and reduces cognitive workload for tubular continuum robot navigation tasks. It is clear that auditory display will not replace visual navigation methods that have been refined and successfully implemented in the surgical routine over many decades. However, auditory display could be employed in many instances during which the view of a patient or situs is compromised or when monitors are inconveniently placed.
Chapter 5

Novel Mixed Reality for Laparoscopy


About this chapter

This chapter introduces mixed-reality navigation for a laparoscopic surgery system using a head-mounted display to integrate the displays from a laparoscope, navigation system, and diagnostic imaging. This provides context-specific information to the surgeon. An immersive auditory feedback was designed to provide the user with 3-D position cues to aid guidance during intraoperative surgical navigation. The results of peg identification and transfer tasks evaluated with sixteen surgeons show that the ergonomics of laparoscopic procedures could be improved while minimizing the necessity of additional monitors in the operating room.
5.1 Novel Mixed Reality for Laparoscopy

For several years now, surgeons have been aware of the greater physical stress and mental strain during minimally invasive surgery (MIS) compared to their experience with open surgery [83, 116]. Limitations of MIS include lack of adequate access to the anatomy, perceptual challenges and poor ergonomics [85]. The laparoscopic view only provides surface visualization of the anatomy. The internal structures are not revealed on white light laparoscopic imaging, preventing visualization of underlying sensitive structures. This limitation could lead to increased minor or major complications. To overcome this problem, the surrounding structures can be extracted from volumetric diagnostic or intraprocedural CT/MRI/C-arm CT imaging and augmented with the laparoscopic view [51, 110, 135]. However, interpreting and fusing the models extracted from volumetric imaging with the laparoscopic images by the surgeon intraoperatively is time-consuming and could add stress to an already challenging procedure. Presenting the information to the surgeon in an intuitive way is key to avoiding information overload for better outcomes [40]. Ergonomics also plays an important role in laparoscopic surgery. It not only improves the performance of the surgeon but also minimizes the physical stress and mental demand [46]. A recent survey of 317 laparoscopic surgeons reported that an astonishing 86.9% of MIS surgeons suffered from physical symptoms of pain or discomfort [112]. Typically, during laparoscopic surgery, the display monitor is placed outside the sterile field at a particular height and distance, which forces the surgeon to work in a direction not in line with the viewing direction. This causes eyestrain and physical discomfort of the neck, shoulders, and upper extremities. Continuous viewing of the images on a monitor can lead to prolonged contraction of the extraocular and ciliary muscles, which can lead to eyestrain [112]. This paper aims to address the problem of improving the image visualization and ergonomics of MIS procedures by taking advantage of advances in the area of virtual, mixed and augmented reality.

5.2 Mixed Reality Navigation System

A novel MRNLS application was developed using the combination of an Oculus Rift Development Kit 2 virtual reality headset, modified to include two front-facing pass-through cameras, navigation system, auditory feedback and a virtual environment created and rendered using the Unity environment.
5.2. Mixed Reality Navigation System

5.2.1 Mixed Reality Head-Mounted Display (HMD)

The Oculus Rift Development Kit 2 (DK2) is a stereoscopic head-mounted virtual reality display that uses a 1920x1080 pixel display (960x1080 pixels per eye) in combination with lenses to produce a stereoscopic image for the user with an approximately 90 degrees horizontal field of view. The headset also features 6 degrees of freedom rotational and positional head tracking achieved via gyroscope, accelerometer, magnetometer, and infrared LEDs with an external infrared camera. A custom fitted mount for the DK2 was designed and created to hold two wide-angle fish-eye lens cameras, as shown in Figure 5.1.
Fig. 2. Software layout of the mixed reality navigation for laparoscopic surgery.

The cameras add the ability to provide a stereoscopic real-world view to the user. The field of view for each camera was set to 90 degrees for this mixed reality application. The double-camera mount prototype was 3D printed allowing for adjusting the interpupillary distance as well as the angle of inclination for convergence between the 2 cameras. These adjustments were designed to be independent of one another. Camera resolution was at 640x480 pixels each. It was found that the interpupillary distance had the greatest contribution to double vision and was hence adjusted differently from one user to another. The prototype was designed to be as lightweight and stable as possible to avoid excessive added weight to the headset and undesired play during head motion respectively. An existing leap motion attachment was used to attach the camera mount to the headset.

5.2.2 Mixed Reality Navigation Software

A virtual environment was created using Unity 3D and rendered to the Oculus Rift headset worn by the user (Figure 5.2). As seen in Figure 5.3, a real-world view provided by the mounted cameras is virtually projected in front of the user. Unlike the real-world view, virtual objects are not tethered to the user’s head movements. The combination of a real-world view and virtual objects creates a mixed reality environment for the user. Multiple virtual monitors are arranged in front of the user displaying a laparoscope camera view, a navigation view, and diagnostic/intraprocedural images.
5.2. Mixed Reality Navigation System

**Diagnostic/intraprocedural images** A custom web server module was created for 3D Slicer allowing for external applications to query and render DICOM image data to the headset. Similar to the VR diagnostic application, we have developed a web server module in 3D Slicer to forward volume slice image data to the MR application, created using the Unity game engine. The Unity application created a scene viewable within the HMD and query the 3D Slicer Web Server module for a snapshot of image slice windows, which is then displayed and arrayed within the Unity scene. The Unity application renders the scene stereoscopically with distortion and chromatic aberration compensating for the DK2’s lenses. At startup, image datasets were arrayed hemispherically at a distance allowing for a quick preview of the image content, but not at the detail required for in-depth examination. Using a foot pedal while placing the visual reticule on the images brings the image window closer to allow for in-depth examination (see accompanying video).

**Surgical navigation module (iNavAMIGO)** The iNavAMIGO module was built using the Wizard work flow using Qt and C++. The advantage of this work flow is that it allows the user to step through the different steps of setting up the navigation system in a systematic method. The Wizard work flow consists of the following steps:

1. Preoperative planning
2. Setting up the OpenIGTLink server and the instruments
3. Calibration of the tool
4. Patient to image registration
5. Setting up displays
6. Postoperative assessment
7. Logging data

**Setting up the OpenIGTLink server and the instruments** In this step, an OpenIGTLink server is initiated to allow for the communication with the EndoTrack module. The EndoTrack module is a command line module that interfaces to the electromagnetic tracking system (Ascension Technologies, Vermont, USA) to track the surgical instruments in real-time. Further an additional server is setup to communicate with a client responsible for the audio feedback. Visualization Toolkit (VTK) [132] models of the grasper and laparoscope are created and set to observe the sensor transforms. Motion of the sensor directly controls the display of the instrument models in 3D Slicer.
Calibration and registration Since the EM sensors are placed at an offset from the instrument tip, calibration algorithms are developed to account for this offset. The calibration of the instruments is performed using a second sensor that computes the offset of the instrument tip from the sensor location. Although the iNavAMIGO module supports a number of algorithms to register the EM to imaging space, in this work we have used fiducial-based landmark registration algorithm to register the motion of the instruments with respect to the imaging space.

Displays The display consists of three panes. The top view shows the 3D view of the instruments and the peg board. This view also displays the distance of the grasper from the target and the orthogonal distance of the grasper from the target. The bottom left view shows the virtual laparoscopic view while the bottom right view shows the three-dimensional view from the tip of the grasper instrument. The instrument-display models and the two bottom views are updated in real-time and displayed to the user. The display of the navigation software is captured using a video capture card (Epiphan DVI2PCIe, Canada) and imported into the Unity game development platform. Using the VideoCapture API in Unity, the video from the navigation software is textured and layered into the Unity Scene. The navigation display pane is placed in front of the user at an elevation angle of -30 degrees within the HMD (Figure 5.3, bottom).

Laparoscopic and Camera View Video input from both front-facing cameras mounted on the HMD was received by the Unity application via USB. The video input was then projected onto a curved plane corresponding to the field of view of the web cams in order to undistort the image. A separate camera view was visible to each eye creating a real-time stereoscopic pass-through view of the real environment from within the virtual environment. Laparoscopic video input was also received by the Unity application via a capture card (Epiphan DVI2PCI2, Canada). The laparoscopic video appears as a texture on an object acting as a virtual monitor. Since the laparoscopy video is the primary imaging modality, this video is displayed on the virtual monitor placed 15 degrees below the eye level at 100cm from the user. The virtual monitor for the laparoscopy video is also be placed directly in line with the hands of the surgeon to minimize the stress on the back, neck and shoulder muscles, see Figure 5.3 (bottom) and videos accompanying the paper.
5.2.3 Audio Navigation System

The auditory feedback changes corresponding to the grasper motion in 3DOFs. In basic terms, up-and-down (elevation) changes are mapped to the pitch of a tone that alternates with a steady tone so that the two pitches can be compared. Changes in left-and-right motion (azimuth) are mapped to the stereo position of the sound output, such that feedback is in both ears when the grasper is centered. Finally, the distance of the tracked grasper to the target is mapped to the inter-onset interval of the tones, such that approaching the target results in a decrease in inter-onset interval; the tones are played faster. The synthesized tone consists of three triangle oscillators, for which the amplitude and frequency ratios are 1, 0.5, 0.3 and 1, 2, and 4, respectively. The frequency of the moving base tone is mapped to changes in elevation. The pitches range from note numbers 48 to 72 on the Musical Instrument Digital Interface (MIDI). These correspond to a frequency range of 130.81 Hz to 523.25 Hz, respectively. Pitches are quantized to a C-major scale. For the y axis (elevation), the frequency $f$ of the moving base tone changes as per the elevation angle. The pitch of the reference tone is MIDI note 60 (261.62 Hz). Thus, the moving tone and reference tone are played in a repeating alternating fashion, so that the user can compare the pitches and manipulate the pitch of the moving tone such that the two pitches are the same and elevation $y=0$. Movement along the azimuth (x-axis) is mapped to the stereo position of the output synthesizer signal. Using this mapping method, the tip of the grasper is imagined as the ‘listener,’ and the target position is the sound source, so that the grasper should be navigated towards the sound source.

5.3 Experimental Methods

A pilot study was conducted to validate the use of the head mounted device based mixed reality surgical navigation environment in the operating room simulated by a Fundamentals of Laparoscopic Surgery [52] skills training box. FDA Institutional Review Board (IRB) approval [147] was waived for this study.

Participants were asked to complete a series of peg transfer tasks on a previously validated FLS skills trainer, the Ethicon TASKit (Train Anywhere Skill Kit, Ethicon Endo-Surgery Cincinnati, OH, USA). Modifications were made to the Ethicon TASKit to incrementally advance the difficulty of the tasks as well as to streamline data acquisition (see Figure 5.4). Two pegboards were placed in the box instead of one to increase the yield of each trial. The pegboards were placed inside a plastic container that was filled
Figure 5.3: (top) User with the MRNLS performing the trial (bottom) view provided to the user through the HMD. Virtual monitors show the laparoscopy view (panel a - red hue) and the navigation system display (panel b, c, d). The surrounding environment (label e) can also be seen through the HMD.
5.3. Experimental Methods

FIGURE 5.4: Example trajectory of the grasper as recorded by the EM sensor

with water, red dye, and cornstarch to simulate decreased visibility for the operator and increased reliance on the navigation system. Depending on the task, visualization and navigation would be performed using laparoscopic navigation with CT imaging (LN-CT, standard of care) or mixed reality navigation (MRNLS).

Tasks 1 and 2 - Peg Transfer Using standardized instructions, participants were briefed on the task goals of transferring all pegs from the bottom six posts to the top six posts and then back to their starting position. This task was done on two pegboards using the LN-CT (task 1) and then repeated using the head mounted device (task 2). No additional information or navigation system was given to the participants while wearing the head mounted device other than the laparoscopic camera feed. To determine time and accuracy of each trial, grasper kinematics were recorded from the grasper sensor readings, including path length, velocity, acceleration, and jerk.

Tasks 3 and 4 - ‘Tumor’ Peg Identification and Transfer Tasks 3 and 4 were designed as a modified peg transfer with a focus on using the navigation system and all information to identify and select a target ‘tumor’ peg from surrounding normal pegs, which were visually similar to the ‘tumor’ peg but distinct on CT images. Participants were instructed to use the given navigation modalities to identify and lift the ‘tumor’ peg on each pegboard and transfer it to the last row at the top of the pegboard. Task 3 had participants use the standard approach of laparoscopy and CT guidance (LN-CT), whereas task 4 was done with the laparoscopic feed, audio navigation, and
Chapter 5. Novel Mixed Reality for Laparoscopy

FIGURE 5.5: Example trajectory of one of the trials, from which the kinematic parameters have been derived.

3D renderings integrated on the mixed reality HMD environment, i.e., the MRNLS. Metrics recorded included time to completion, peg drops, incorrect peg selections, and probe kinematics such as path length, velocity, acceleration, and jerk.

Tasks 5 and 6 - ‘Tumor’ Peg Identification and Transfer Through Sensitive Structures  For the final two tasks, modifications were made to the laparoscopic skills trainer box to stress the navigation system and recreate possible intraoperative obstacles such as vasculature, nerves, and ducts. Using a plastic frame and conductive wire, an intricate structure was made that could easily be attached for tasks 5 and 6. The structure held the conductive wire above the pegboards in three random, linear tiers (Figure 5.4). A data acquisition card (Sensoray S826, OR, USA) was used to asynchronously detect contact with the wires by polling the digital input ports at a sampling rate of 22Hz. Contact between the grasper and the wires could then be registered and tracked over time. Operators were asked to identify the radiolabeled ‘tumor’ peg and transfer this peg to the last row on the pegboards. However, in this task they were also instructed to do so while minimizing contact with the sensitive structures. In task 5, participants used the current standard approach of LN-CT, while in task 6, they used the proposed MRNLS system with fully integrated audio feedback, 3D render-based, and image guided navigation environment viewed on the HMD.
5.4 Participants  A total of 16 surgeons with different experience levels in laparoscopic surgery volunteered to participate in the study and were assigned to novice or experienced subject groups. Novice surgical subjects included participants who performed more than 10 laparoscopic surgeries as the secondary operator but less than 100 laparoscopic surgeries as the primary operator. Experienced subjects were those who performed more than 100 laparoscopic surgeries as primary operator.

Questionnaire and Training Period  Following each task, participants were asked to complete a NASA Task Load Index questionnaire to assess the workload of that approach on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration.

Statistical Analysis  The Wilcoxon signed-rank test for non-parametric analysis of paired sample data was used to compare the distributions of metrics for all participants by task. The Mann-Whitney U test was used to compare distributions in all metrics between novice and expert cohorts, and \( p < .05 \) was considered statistically significant.

5.4 Results and Discussion

Figure 5.5 shows an example trajectory of one of the trials, from which the kinematic parameters have been derived. Please see accompanying videos for trial details.

Tasks 1 and 2  On the initial baseline peg transfer task with no additional navigational modalities, participants took longer to complete the task when viewing the laparoscopic video feed on the mixed reality HMD, as part of the MRNLS (standard: 166.9s; mixed reality: 210.1s; \( p = 0.001 \)). On cohort analysis, expert participants showed higher significance in time to completion than novices (\( p = 0.004, \ p = 0.011 \)). Additionally, there was no difference in number of peg drops or kinematic parameters such as the mean velocity, mean acceleration, and mean jerks per subject amongst all participants or by expertise. During these baseline tasks, mental demand, physical demand, and frustration were significantly increased (\( p < 0.05 \)) when using the mixed reality HMD environment with mildly significant decrease in perceived performance (\( p = 0.01 \)). However, effort and temporal demand showed no significant differences amongst all subjects nor novices and experts alone.
Tasks 3 and 4  Compared to the standard LN-CT in task 3, all participants showed significant decrease in time to completion with the aid of the MRNLS (decrease in time = -20.03 s, \( p = 0.017 \)). When comparing the addition of the MRNLS in task 4 to the standard approach in novice and expert participants, novice participants showed significant improvements in mean velocity, mean acceleration, and mean jerks between tasks 3 and 4, compared to only mean velocity in experts. Mental demand was significantly decreased when combining the results of both novice and expert participants (\( p = 0.022 \)) and there was near significance for performance (\( p = 0.063 \)) and effort (\( p = 0.0089 \)) for the MRNLS.

Tasks 5 and 6  Tasks 5 and 6 were designed to compare the standard LN-CT and proposed MRNLS on a complex, modified task. These final tasks again demonstrated significantly faster time to completion when using the MRNLS in task 8 (100.74 s) versus the LN-CT in task 7 (131.92 s; \( p = 0.044 \)). All other kinematic metrics such as average velocity, acceleration, jerks, as well as time in contact with sensitive wire structures, peg drops, or incorrect selections showed no significant difference between navigation modalities for all participants, novices, or experts. Amongst novice participants, there was a decrease in the means of time to completion (-45.5 s), time in contact (-14.5 s), and path length (-432.5 mm) while amongst experts there was a smaller decrease in these metrics (-20.1 s, 2.12 s, -163.1 mm) for the MRNLS. Novices were twice as likely to make an incorrect selection using LN-CT versus MRNLS, however, and experts were 3 times as likely. According to the NASA Task Load Index values, the effort that participants reported to complete the task was significantly lower using the MRNLS compared to the LN-CT (Difference of -1.375, \( p = 0.011 \)). Upon analysis by expert group, this significance is present among the novice participants but not among expert participants (Novices: -2.57, \( p = 0.031 \); Experts: -0.44; \( p = 0.034 \)). There was a similar result for frustration that was near significance (All participants: -1.38, \( p = 0.051 \); Novices: -2.43, \( p = 0.063 \); Experts: -0.22, \( p = 1 \)).

5.5 Conclusion

We have validated the use of a novel mixed reality head mounted display navigation environment for the intraoperative surgical navigation use. Although further studies are warranted, we find the use of this novel surgical navigation environment proves ready for in-vivo trials with the objective of additionally showing added benefits with respect to surgical success, complication rates, and patient-reported outcomes.
Chapter 6

Resection Guidance for Open Liver Surgery


About this chapter

Surgeons who use image guidance during open liver surgery are restricted by having to continuously switch between viewing the navigation system screen and the patient during an operation. As an alternative mode of interaction with navigation systems for open liver surgery, this chapter presents an auditory display system for open liver surgery for guiding the tracked instrument towards and remaining on a predefined resection line. Qualitative results from the user study show that the proposed auditory display is recognized as a useful addition to the current visual mode of interaction, reducing time looking on the screen and increasing accuracy for resection guidance was significantly improved when using auditory display as an additional information channel, however, the overall time for the resection task was shorter without auditory display. By reducing dependence on the visual modality during resection guidance, the auditory display is well suited to become integrated in navigation systems for liver surgery.
6.1 Introduction

Surgical navigation systems provide intraoperative assistance by displaying surgical instruments in relation to preoperative liver planning data [95, 117]. Similar technology is already exhibiting good results in orthopedic surgery and neurosurgery.

The navigated resection of tumors in the liver enables a higher degree of precision and could lead to smaller resection volumes, more frequent tumor-free margins, and thus, improved outcomes [111]. However, surgeons frequently need to consult the navigation system, which leads to increased mental load and time pressure during surgery [101].

Viewing the navigation system’s screen interrupts the surgeon’s focus on the working area. The purpose of this work is to integrate concepts from the field of auditory display to enhance intraoperative visualizations for navigated liver surgery. Auditory display is applied as an additional information channel. It might reduce the dependence on the visual display. Thereby, it frees the eyes for other tasks, and provides high temporal resolution and fast processing [80]. Although auditory alarms and monitoring devices are commonplace in operating rooms, audio has been a neglected modality in navigated surgery. During liver resection, the auditory channel is primarily unused by the surgeon, and thus is more open to provide information that supplements or complements the visual display.

This work presents a solution for a primary problem in navigated liver surgery: the excessive dependence on the visual display during resection guidance. This work focuses on marking the resection path on the liver surface. The resection marks help determine the subsequent cut path through the liver parenchyma. Until now, resection guidance has been supported solely by visual displays.

6.1.1 Related Work

Despite research in the field of alarms and auditory status monitoring [129, 153], only a few publications describe auditory display to support navigated surgical tasks. To the knowledge of the authors, only two groups have developed auditory displays for this field. Wegner et al. [155, 156] introduced a navigation system for the blind placement of a biopsy needle. Based on an electromagnetic tracked needle, surgeons attempted to follow a planned trajectory inside a custom gelatin phantom with an embedded target (tumor mimic). Three relevant methods are described in their papers: The first method applies a beat interference method, for which two sinus waves (one with fixed, the other with varying frequency) are mapped to the positional
error in one spatial dimension. By bringing the frequencies in harmony, the user can guide the needle. However, to guide the needle in 3D, three wave pairs are needed which can be misleading if the sinus waves of one spatial dimension are in harmony with the sinus wave of another dimension. The second method is used to perform intraoperative distance measurements. Therefore, the user determines a start point with the tracked needle. As the user moves the needle away from the start point, short sound events are triggered in equidistant distances. For example, at each millimeter an audible click is triggered and at each centimeter a speech sample that indicates the distance in centimeter is triggered. The third method aims to explore surface characteristics of anatomical 3D models. The angle between the needle and the surface of the 3D models is measured. Based on a 2D wave table, a homogeneous tone is produced when the needle is located in a homogeneous surface region. The sound frequency varies when the user passes the needle across an uneven surface. These changes in sound frequency are intended to guide the operator. Although the proposed methods by Wegner et al. showed great potential, formal evaluations were not reported in literature.

A group from the University Medical Center in Utrecht, Netherlands presented a method for auditory display for neurosurgery. In this case, auditory display was applied to resect a floral foam phantom meant to simulate brain tissue [157] and later for operation on six patients [158]. When the tracked instrument approached the tumor, the system produced a pure tone with a duration of 0.1 seconds emitted approximately 3 times per second. Frequency and amplitude of the tone increased when approaching the target. Directional information was not encoded. The task involving the phantom consisted of removing a target volume from a block of floral foam. The resection using auditory display was compared to resection with visual-based navigation (using a computer monitor for display) and with a heads-up display (for which contours of segmented structures were superimposed onto real-time microscopic images) [157]. Resection quality (similarity of actual and planned resection volume) and resection time did not significantly differ when comparing the results between interaction with auditory display and visual display. In clinical tests, the auditory display was employed on 6 patients and evaluated on instrument handling during the course of surgery. The study results showed that the speed of the instrument tip was not significantly faster when using auditory support. However, in both studies participants felt that they performed better when auditory feedback was added because they could focus more on the working area.
6.2 Materials and Methods

This section describes an auditory display for image-guided liver surgery. First, the existing navigation system, the underlying planning data, and the auditory feedback engine are described. Second, a novel method for auditory resection guidance is introduced. Third, the performed evaluation of the developed auditory display is described.

6.2.1 Navigation System: CAS-One

To assist surgeons placing or moving surgical instruments, surgical navigation systems are applied. Generally based on visual display, the position of tracked surgical instruments such as dissectors, ultrasound probes, and needles, can be viewed in relation to 3D planning models. The planning models are generated from CT or MRI scans acquired from the patient’s liver using planning software for liver surgery [23]. Besides anatomical 3D models of the liver, the planning model includes relevant functional information such as a virtual resection surface [91]. One element of the virtual resection surface is the resection line (on the organ surface), which is used within the evaluation of this work. The CAS-One surgical navigation system (CAScination AG, Switzerland) is used within this project to visualize planning models relative to tracked surgical instruments. CAS-One provides visual support during liver resection and has been evaluated in the context of clinical trials [117]. The system consists of an infrared-based optical tracking system (Vicra, Northern Digital Inc., Canada), an ultrasound device, a touchscreen, and a computer unit. A rigid body with passive markers is intraoperatively mounted on surgical instruments (ultrasound and dissector), or on a pointer for laboratory experiments. Landmark-based registration is applied during the first phase of the procedure, so that the modalities of the physical patient and 3D planning model align.

6.2.2 The Auditory Feedback Engine

Two development platforms were chosen, MeVisLab [62] for medical image processing framework and SuperCollider [104] for audio synthesis. The developed prototype application was independently implemented from the certified navigation software. All necessary visualizations for navigation support, including the representation of tracked instruments and 3D planning models were provided by a customized MeVisLab application installed
on the navigation system. In addition, the application continuously sends distances and information about associated object entities to the SuperCollider engine using the OSC (Open Sound Control) protocol. In the following subsection, a novel method for auditory resection guidance is presented.

### 6.2.3 Auditory Resection Guidance

For complex liver operations, resection surfaces are preoperatively defined using surgical planning software. Intraoperatively, surgeons aim to resect according to these plans with the help of surgical navigation systems. Before cutting the liver, the resection path is marked on the organ surface. An accurate marking of the resection path is a decisive factor for the quality of the resection as a whole. The marks provide orientation aid during the relatively long cutting phase (1 to 2 hours) and thus impact the subsequent cut path through the liver parenchyma.

The proposed auditory feedback for resection guidance is a function of $\delta$, where $\delta$ describes the distance between surgical instrument tip and the nearest point on the planned resection line. A precomputed Euclidean distance transformation provides $\delta$ for each position in the patient dataset. Thus, the auditory feedback engine can access $\delta$ in constant computation time.

The distance from the instrument tip on either side of the resection line is divided into safe, warning, and outside margins:

The first margin, safe, is located on both sides of the resection line. When the instrument tip is in this margin, it is permissible for the surgeon to resect or mark. The second margin, warning, is located outside of the safe margin and reaches from the outer border of the safe range to the warning-range width. When the instrument tip is located in this margin, the surgeon is roughly near the planned resection line but not close enough for a safe resection marking. Finally, the outside margin includes all distances that are further from the resection line than the outside width of the warning margin. The distances for the safe and warning margins can be set by the surgeon depending on the surgical situation.

To communicate the presence of these different margins and to direct the surgeon towards the resection line, two tones for auditory resection guidance are proposed which correspond to both the safe and warning margins. When the instrument tip is in the safe margin ($0 \leq \delta \leq \text{safe width}$), signifying that it is permissible to mark the resection line, the safe tone is produced. When the tip is in the warning range (safe width $< \delta \leq \text{warning width}$), the warning tone aids the surgeon in directing the instrument tip to enter the in-range
margin. When the instrument tip is in the outside margin (warning width < $\delta$), no tone is produced. The combination of both warning and confirmation tones provides the user with a way of both locating and remaining on the planned resection line.

**Prior Work**

Based on our prior work in the field of auditory display [11, 13] a refined resection guidance model is introduced. Within this prior work, different auditory display models for resection guidance have been developed. These first models ranged from very simple, using only a single modulated sine wave, to more complex, featuring banks of digital resonating bodies, to ones that modeled real-life sounds such as ringing bells. A preliminary user study was carried out, first to provide a general consensus on whether or not auditory display would be beneficial for navigated liver surgery, and then to gather clinicians’ opinions regarding the functions and aesthetics of a range of exploratory auditory display models. The study revealed that auditory display was recognized by a majority of surgeons to be a useful addition to the visual interaction with the navigation system. A systematic analysis of user comments confirmed that the auditory feedback also combated main limitations of the current navigation system, i.e., the dependence on the visual display and the lack of notification when approaching risk structures. In addition, the meaning of the sound and the interaction with the system were easily learned within less than two minutes of training. The second aim of these preliminary studies was to gather comments about a selection of different auditory display models for resection guidance. Comments regarding these first models were used in the development of the refined resection guidance model.

**A Refined Resection Guidance Model**

As part of the process of continual evaluation and refinement, the most popular elements of the exploratory resection guidance sound models were used to create a single model that could be used for the quantitative evaluation described in this paper. The concept of this model is broadly similar to a Geiger counter in that the inter-onset interval of a series of very short auditory events is one parameter that is mapped to the urgency of a situation. In the case of resection guidance, an increase in distance from the planned resection line causes an increase in urgency.
6.2. Materials and Methods

**Safe Margin Tone:** When the instrument tip is in the safe margin, a ‘safe margin tone’ is produced to guide the surgeon to \( \delta = 0 \) and to inform that it permissible to mark the resection at that point. The safe margin tone consists of two elements. First, a two-pole resonant filter with a frequency of 698.5 Hz (corresponding to MIDI note 77) and a 60 dB decay time of 1.0 second is triggered repeatedly at a variable inter-onset interval. At a distance of \( \delta = 0 \), this inter-onset interval is 0.66 seconds, and at the edge corresponding to 698.5 Hz (corresponding to MIDI note 77) of the safe margin, the interval is 0.18 seconds. By moving the instrument tip so that the inter-onset interval is longer, the surgeon is guided towards the planned resection line. Second, a bank of sine oscillators with frequencies of 220.0, 261.6, 349.2, 440.0, and 523.3 Hz (corresponding to MIDI notes 57, 60, 65, 69, and 72) is produced when \( \delta = 0 \) and muted at all other distances. Thus, this confirmation element informs the surgeon when it is optimal to mark the line.

**Warning Margin Tone:** When the instrument tip is outside the safe margin, a warning margin tone helps guide the surgeon towards the safe margin. Similar to the safe margin tone, a series of repeating two-pole resonant filters are employed. In this model, three primary auditory characteristics are used to relay distance information and thus convey a sense of urgency (see [44]) when approaching the outside of the warning margin and a sense of calm when approaching the safe margin. Each of these variables varies linearly over \( \delta \).

**Inter-Onset Interval:** The inter-onset interval times of the triggered resonators varies with distance. At the inside edge of the warning margin, the interval is 0.33 seconds, and at the outside edge, the interval is 0.09 seconds.

**Tone Length:** The 60 dB decay time of the resonating tones varies with distance. At the inside edge of the warning margin, the decay time is 1.0 second, and at the outside edge, the decay time is 0.5 seconds.

**Pitch:** The pitches of the resonators relay both distance and the side of the resection line in which the instrument tip is located. Moving the tracked instrument to the left of the resection line causes the tones to rise in pitch. Directly at the left inside edge of the warning margin, the pitch of the resonator is 784.0 Hz and over the width of the warning margin increases to 880.0, 1046.5, 1174.7, 1396.9, 1567.9, and 1760 Hz (corresponding to MIDI notes 79, 81, 84, 86, 89, 91, and 93). Moving to the right of the resection line causes a fall in pitch. Directly at the right inside edge of the warning margin, the
pitch of the resonator is 587.3 Hz and over the width of the warning margin decreases to 523.3, 440.0, 392.0, 349.2, 293.7, and 291.6 Hz (corresponding to MIDI notes 74, 72, 69, 67, 65, 62, and 60). The combination of onset frequency, tone length, and rising and falling pitch relay to the surgeon both the distance to the resection line and which side of the resection line the tracked instrument occupies. By moving the instrument back and forth across the safe and warning margins, the surgeon should be able to use these auditory cues to place the instrument tip directly on the planned resection line and be aware when the instrument deviates from the optimal planned line.

6.2.4 Evaluation

The objective of this work is to evaluate the suitability of auditory display as an enhancement to visual display in navigated liver surgery. Therefore, a clinically oriented study was conducted in which participants were asked to accomplish surgical tasks on a special manufactured 3D liver phantom. User performance was evaluated quantitatively by analyzing instrument movement and aspects of human-computer interaction.

Liver Phantom

Because an evaluation of the new interaction techniques would not be fruitful unless a physical representation of a human liver was present, a CT liver dataset with segmented anatomical structures was used to create a stereolithographical model of the liver. This liver phantom only includes an outer shell with a cuboid cavity in the front to enable access to the interior with the tracked instrument. A piece of floral foam (Colorfoam, mosy GmbH, Thedinghausen, Germany) was trimmed to fit inside this cavity to provide tactile resistance to the tracked instrument during testing (Figure 6.1). The phantom was mounted on a wooden board together with a trackable marker shield, which allowed for easy calibration with the navigation system’s camera.

Reference Criteria

Appropriate reference criteria for the evaluation were found by an analysis of which characteristics of the auditory display could potentially improve the surgical workflow of open liver interventions. The dependence on the visual display, which is a primary problem, was observed during the study. In addition, the movement of the instrument tip was compared.
The following reference criteria were defined:

- Total time participants looked at screen relative to looking at phantom
- Time needed to draw the resection line
- Mean distance between planned and subject-drawn resection lines

### 6.2.5 Experiment Design

The tests were performed in a clinical environment with 12 surgeons from the Robert-Bosch Hospital in Stuttgart, Germany. The liver phantom was affixed to an examination table. Participants were asked to stand next to the liver phantom. The navigation systems display was placed on the opposite side of the table, as common during clinical interventions that employ video displays (see Figure 6.2).

First, two training tasks had to be performed by the participants. Therefore, a red resection line and a 3D model of the liver phantom were displayed on the navigation system’s display. Participants were asked to transfer the resection line visualized on the video monitor to the liver phantom by lightly marking the path on the floral foam using a pointer whose position was tracked by the navigation system. Participants were instructed to look at the video monitor only if necessary and to concentrate on the surgical instrument and the liver phantom. One training task was to be performed without auditory support (visual only) and the other training task with auditory support (combined condition). Second, eight test tasks had to be performed by the participants. Thereby, four resection lines (with the 3D model of the liver phantom) were separately presented two times on the screen of the navigation system. The test tasks were divided into two groups: While four tasks (V1 to V4) did not make use of auditory display (visual only), the other four
The tests were performed in a clinical environment in Stuttgart, Germany. The liver phantom was presented to the participants as a way to simulate a real surgical environment. The phantom consisted of floral foam inside a liver phantom. This shows the surgeon observing the liver phantom, which prohibited two test tasks with the same resection line to be performed in succession.

**Figure 6.2:** Setup of the navigation system and the liver phantom within the user study. The participating surgeon uses the tracked instrument to mark the planned resection line on the liver phantom. This shows the surgeon observing the liver phantom.

**Figure 6.3:** Surgeon glancing at the navigation system screen

Tasks (A1 to A4) presented both visual and auditory displays (combined condition). The presentation of these test tasks occurred in random order. In order to reduce the memory effect, a randomization algorithm was chosen which prohibited two test tasks with the same resection line to be performed in succession.
Before starting each test, participants were informed of the group to which the upcoming test belonged. Participants were further asked to look at the video monitor only when necessary. After marking the resection line on the foam, participants were asked to draw the final resection line a second time. The position and orientation of the pointer were recorded during the whole experiment. In addition, all tests were recorded with a video camera. Third, each participant completed a questionnaire consisting of personal questions (regarding age, gender, handedness, surgical experience, etc.) and statements for which the participants chose the degree to which they agreed or disagreed.

Analysis

A post-experiment video analysis provided the timing data for subjects’ looks at the screen or the phantom, respectively. These data were used to construct the time budget data and the mean times between changes of looks between phantom and screen. Data for marking accuracy was sampled using the tracking mechanisms of the navigation system. The distance \( \delta \) between the pointer tip position and the nearest point on the resection line was calculated for each sample. Because these samples were taken in equal temporal distance, arithmetic means of \( \delta \) would be skewed by different drawing speeds. When, for example, a subject did not move the tracked pointer at all, remaining in a position with low \( \delta \), the data with low \( \delta \) would accumulate over time and thus over-represent a low \( \delta \) in the mean. Therefore, the mean distance between planned resection line \( I_p \) and subject-drawn resection line \( I_d \) was introduced as a benchmark for marking quality. Given the arclength \( s \) on \( I_d \), the mean distance is defined in the continuum as

\[
\frac{1}{L_d} \int_{I_d} \delta(s) \, ds
\]

where \( \int_{I_d} ds \) is the total length of \( I_d \).

Statistical analysis consisted of comparisons of the arithmetic means between data for the auditory and visual conditions. We tested differences using the Wilcoxon test and defined \( p < 0.05 \) as statistically significant.
6.3 Results

For all reference criteria, we found differences between the visual and combined conditions. For the visual only condition, subjects looked nearly 100% of the time at the screen. This time was reduced to around 10% for the condition where the auditory signal was additionally available (Figure 6.4).

When the additional auditory information was available to the subjects, deviation from the planned resection line was smaller than with the visual information alone (Figure 36.4. However, subjects could not mark the resection line as fast as when relying on the auditory information as opposed to the visual display only (Figure 6.5). Qualitative data for this evaluation was gathered from both voluntary comments from the participants as well as a questionnaire filled out by each participant. From this data, the direction of future development may be guided. Table 6.1 lists each statement and the average response and standard deviation.

![Figure 6.4: Proportion of time looking at screen between planned resection line and subject-drawn resection line for combined and visual condition](image)

6.4 Discussion

The clinically oriented tests performed in the context of this work revealed that auditory feedback could be a beneficial extension to surgical navigation systems. However, tests during surgical interventions have to still be performed in order to further adapt the proposed sound model for clinical routine. Specifically, the effects of environmental sounds such as speech, cutting
device noise, or warning signals from anesthesia devices on the surgeons’ interaction and recognition of the auditory display must be evaluated. Overall, the qualitative questionnaire showed that the subjects found the approach of using auditory display for resection marking as promising and well suited for the task. Surgeons requested aesthetically pleasing sounds that could be tolerated to hear over a time range long enough for surgical tasks. However, it is difficult to discover sound models that meet these requirements as well as providing the required resection guidance support in a clear way. An adequate variety of sound models might be provided to give surgeons a choice of equally functional sound models. In addition, these sound models might be made easily configurable for surgeons or assistants to improve acceptance and reduce possible annoyance. Moreover, industrial standards such as IEC 60601-1-8 for alarms in medical equipment [75] have to be heeded to meet legal requirements for audio in operating environments. In our evaluation, the auditory display significantly reduced the time that surgeons looked on the navigation system screen. Simultaneously, the auditory display reduced the mean distance to the planned resection line. However, the mean time needed for each task was almost doubled. Thus, our auditory display reduces the dependency on the visual display and increases accuracy, albeit at the cost of task completion time. One contributing factor for this increase in task completion time could be that the auditory display alerted the surgeon when leaving the safe margin. This auditory notification may have caused a more cautious (but also more accurate) marking of the resection line than without having this information. A second contributing factor could be the
new introduction of the auditory display to the subjects, many with experience working with visual navigation systems but none with auditory displays. When using the visual display alone, the interaction with the system was similar to using a computer mouse, because the position and movements of the on-screen instrument directly corresponded to those of the physically tracked instrument. Due to the novelty of our method, a longer and more intense training period would likely reduce the task completion time. Conversations with test surgeons revealed that longer task times are not necessarily a negative effect if these longer times accompany an increase in concentration. In the qualitative questionnaire, most subjects agreed that the auditory display helped to concentrate on the resection task. A third contributing factor to increased task completion time could be the study conditions, which did not exactly match the situation of an actual liver resection. Although the task to mark a resection line on a liver phantom is similar to resection marking on a real human liver, one important difference is that during real liver surgery, it would be irresponsible to mark a resection line on the liver by looking nearly 100% of the time at the screen, as several subjects did during our tests. It could be assumed that during actual liver surgery, the time looking at the screen decreases, while the time to mark a resection line increases. For further laboratory studies, the study design should consider conditions that stimulate participants to look more often at the liver phantom, e.g., by showing essential information on the phantom that needs to be perceived by the participant to pass the test. Our method builds upon work by Wegner et al. [156] and Woerdeman et al. [158], who provided groundwork for applying auditory display to navigated surgery tasks. However, our work aims to go beyond basic approaches, such as those of Woerdeman et al., and more deeply explore the ability of alternative modalities like auditory display can
have to reduce stress and increase concentration and accuracy in navigated surgery. Because the auditory display system has shown promise in helping to improve the accuracy of resection marking and reducing the amount of time looking on the screen, it may also improve other aspects of navigated liver surgery, and by extension, other types of navigated surgery as well. In the domain of liver surgery, auditory display has also been explored [11] for notifying the surgeon of potential risk structures in the vicinity of the tracked instrument, such as veins and tumors that should not be damaged. By emitting warning signals when the instrument tip enters predefined distances to the risk structures, the surgeon could be notified of presence of such a risk without having to look at the visual display and taking concentration away from the patient. However, when considering clinical applicability of available navigation systems for liver surgery, it is important to mention that (to the knowledge of the authors) currently, only the marking step is supported with adequate registration accuracy. An accurate, continuous tracking of liver movement during the cutting phase is still under research [8, 33, 149]. The method proposed in this paper intends to support the marking step, but could also be applied for the cutting phase in the future. In addition to navigated liver surgery, auditory display methods could be useful additions to other types of surgery for which visual contact to the situs is important or for which there is a strong dependence on a visual display, including neurosurgery, laparoscopic surgery, and radio-frequency ablation. Further work must discover for which cases auditory display is beneficial and the optimal means of implementing such displays.
Chapter 7

Fluorescence-Guided Open Brain Tumor Surgery


About this chapter

During neurosurgery, protoporphyrin (PpIX) fluorescence allows discrimination of tumor and normal brain tissue during neurosurgery. A handheld fluorescence probe can be used for spectroscopic measurement of 5-ALA-induced PpIX to enable objective detection compared to visual evaluation of fluorescence. However, this requires that the surgeon either view the measured values on a screen or employ an assistant to verbally relay the values. This chapter describes an auditory feedback system for communicating measured fluorescence intensity values directly to the surgeon. The employed auditory display was shown to be intuitive, easy to learn and remember, fast to recognize, and accurate in providing users with measurements of fluorescence intensity or error signal.
7.1 Introduction

Fluorescence imaging based on 5-aminolevulinic acid (5-ALA) visualized using fluorescence-guided resection (FGR) surgical microscopes is an optical guidance system that has been introduced during the past decade for routine clinical application [142]. 5-ALA is a photosensitizer administered prior to the surgical procedure, which is metabolized and accumulated as protoporphyrin (PpIX) in the tumor cells. When excited by light, the tumor re-emits fluorescence of PpIX, thereby enhancing visibility to the surgeon.

The spectroscopic measurement technique applied using a fiber optic probe is based on measurement of a fluorescence spectrum, which is usually displayed on a screen in the OR. A handheld fluorescence (HHF) probe for spectroscopic measurement techniques has been developed at Linköping University. This HHF probe has been evaluated in over 50 patients in the OR for brain tumor resection guidance as a stand-alone system [63, 65], and in combination with both a neuronavigation system [122] and an FGR microscope [123]. When operating with the FGR microscope, the surgeon observes both the surgical site and the fluorescence without requiring additional feedback support. The fluorescence seen through the FGR microscope is conventionally grouped into negative, weak, and strong [123, 142]. This categorization supports the neurosurgeon in decision making on tissue removal. However, for HHF probe measurements, the surgeon cannot reliably perceive the fluorescence signals, specifically weak signals, through vision.
alone. Moreover, several complications during the operation can disturb the measurement, including blood interference and, to a lesser extent, a surgical microscope’s white light lamp or system failure, including misplacement of the probe. In these cases, the surgeon should be informed to correct for the measurement error. Currently, intraoperative fluorescence measurements from the HHF probe are provided on a computer screen and interpreted by an engineer responsible for the system, who verbally relays signal values in the OR (Fig. 7.1). Examples of fluorescence signals measured during surgery are shown in Fig. 7.2. Ideally, the surgeon should be able to receive information from the system without the need for an interpreter while keeping the visual focus on the surgical site. The responsibility of signal interpretation cannot be placed on the surgeons or their assistants; therefore, additional visual or auditory support for acknowledging measurement results could facilitate this process.

Utsuki et al. [148] reported a system which triggered an audible tone when the measured PpIX fluorescence intensity between 632 and 636 nm exceeded a certain level. Neither the sound characteristics nor intensity varied with the changes in fluorescence intensity and did not consider errors that occur during intraoperative measurements. Further reports on audible systems for optical measurements have been limited. To the authors’ knowledge, this paper presents the first and only evaluation of auditory feedback to relay fluorescence intensity values of an HHF probe.
Using sound to transmit changes in data, termed auditory display, has recently gained attention for a small but varied array of clinical applications to aid surgeons during image guidance. Examples in the literature describe auditory display as a means of information retrieval that goes beyond monitoring tasks (such as those used in anesthesia [129]) and helps deliver important navigation information to the clinician to reduce reliance on computer screens or to enhance awareness of important anatomical risk structures in the vicinity of the surgical instrument. Previous implementations of auditory display for image-guided interventions include neurosurgical volume resection [157, 158], temporal bone drilling [150], cochlear implantation [36], liver resection path marking [67], and ablation and biopsy needle placement [18, 22]. Advantages found in previous attempts include heightened awareness of or distance to anatomical risk structures [36, 141, 150], reduced surgical complication rate [141], increased visual focus on the surgical site [67], and improved placement accuracy [18] when using auditory display as either to replace or to augment existing visual support systems. For a review of applications of auditory display in image-guided interventions, see [16].

The aim of this study was to develop and evaluate an auditory display to support fluorescence-guided open brain tumor surgery using an HHF probe in the laboratory based on previously determined clinical fluorescence intensity levels. The investigation is a first evaluation of auditory display to support HHF probe fluorescence measurements and can be implemented into similar optical systems. By using auditory display to support fluorescence-guided brain tumor surgery, the surgeon should be able to “hear” fluorescence values without relying on a computer screen, thereby enhancing visual focus on the surgical site. In addition, auditory feedback should reduce the need for a surgical assistant to verbally relay intensity values to the surgeon, thus receiving values more quickly and reducing missed values due to interpersonal miscommunication.

### 7.2 Methods

#### 7.2.1 Experimental Design

The principles of intraoperative fluorescence measurements and fluorescence quantification are described in previous work [63, 65, 123]. The fluorescence signal levels in this study were replicated in a tray of optical phantoms (Fig. 7.3) to be comparable to the levels measured in the OR [63, 123]. The instrumentation was selected for the study so that all hardware components were compatible with the LabVIEW® software (National Instruments, Inc., Austin, TX) and Open Sound Control (OSC) [160] protocol.
Brain Tumor Phantoms

Four sets of liquid phantoms were prepared to model the actual clinical measurement situation (see Table 7.1). These included zero signal, low signal, high signal, and error signal. The phantoms modeled the optical properties of the brain tumor using ink and intralipid 20% (Fresenius Kabi, Uppsala, Sweden) [64] including tissue autofluorescence (AF) by adding turmeric dissolved in ethanol (zero signal). In two phantom sets (low and high signal), 10 and 30 $\mu$g/$\mu$l of PpIX disodium salt (MP Biomedicals, Illkirch, France) was added to model the low and high fluorescence signals, respectively. The PpIX concentration was chosen to be greater than what is measured in the brain to account for photobleaching effects on the signals and thus avoid variation in the generated sound on one spot. The maximum PpIX peak in the phantoms was at 634 ± 4 nm due to the chemical environment. In a fourth phantom set, the AF was blocked by additional ink to reflect the situation in which the measurements are obstructed by blood or no signal is recorded (error signal). The tray had 96 wells each of 7 mm diameter and 1 cm depth, see Figure 7.3.

Hardware and Signal Analysis Setup

A 405-nm laser (Oxxius) in continuous mode was used to excite the fluorescence, together with an AvaSpec-ULS2048L-USB2 spectrometer (Avantes BV, Apeldoorn, Netherlands) for detection of fluorescence, measuring wavelengths 580 and 1100 nm. A fiber optic probe (Avantes BV, Apeldoorn, Netherlands) was used to measure the fluorescence in the phantoms. The
probe included one central fiber for excitation and six surrounding fibers for fluorescence collection. The total diameter of the fiber bundle was 1.2 mm. A custom spectrometer interface was developed in LabVIEW. The OSC library was embedded in the program to send measurement values to the sound synthesizer. The spectrometer integration time was set to 800 ms to reflect the settings used in the OR.

The intensity from the 600 nm wavelength representing AF and 630 nm wavelength representing PpIX fluorescence was extracted for the analysis. The intensity of the AF at 600 nm was first analyzed to determine whether an error was present. If the intensity was lower than a certain threshold, a signal level was generated at a very large value out of the fluorescence range, in this case 5000. The threshold was set to the intensity of AF at 600 nm measured on the error phantoms after phantom preparation and the average on the zero phantoms. If the intensity was lower than the threshold, the error tone was generated. If the intensity was higher than the upper threshold, the value was set to a constant within the thresholds’ range. This loop was added to compensate for the effect of AF from the well side walls. If no error was identified, quotient of fluorescence intensity at 630 nm and 600 nm was calculated and sent as a “pulse” to the sound synthesizer. The set of threshold quotients for the conversion of these values to the various tones are included in Table 7.1. The principle of fluorescence signal conversion to

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>E</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>E</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>0</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7.4**: The phantom arrangement with three sets of 32 wells, where the content of the wells is indicated by 0 (zero signal), 1 (low signal), 2 (high signal), or E (error signal). The sequence followed by the participants is shown with arrows.
sound is shown as a flowchart in Fig. 7.5. The program required that the values first be calculated after the signal was measured, creating a delay of one pulse. The tone was played back instantly after a new intensity value was sent to the synthesizer.

### 7.2.2 Auditory Display Design

An initial experimental study was conducted with one neurosurgeon well acquainted with the measurement system and clinical application. Various synthesis methods were presented to map the intensity of the fluorescence signal to parameters of the auditory display. These included a continuous mapping of intensity to vibrato (frequency modulation) rates, continuous mapping of intensity to the pitches of two alternating tones for comparison,
and mapping intensity into discrete values for the playback of individual, short composed musical note sequences to represent desired functions, so-called earcons [21]. After the initial study, discrete values were selected as the optimum mapping method, as the categorization of intensity mapped to a selection of a small number of earcons was found to be most appropriate for the clinical task. Continuous auditory displays, while beneficial for auditory display for surgical trajectory navigation in general [67], are harder to translate into quantitative values, in which case a classification-based approach is suggested. This has, for instance, been successfully employed for auditory display for awareness of risk margins in image-guided interventions [36]. Thus, a set of four tones were produced which transmitted one of three intensity levels (zero, low, and high signal) or the error signal. The following four tones were produced:

**Zero Signal** The signal was generated when the fluorescence intensity quotient was 0 to 0.9. The zero signal tone informs the user that the signals were measured correctly but no PpIX fluorescence (tumor) was detected. This was synthesized as a cluster of three sine wave generators with frequencies of 233, 277, and 370 Hz which were played back with an amplitude envelope of 200 ms attack phase, 400 ms sustain phase, and 200 ms release phase. The resulting tone was a calm major G chord \(^1\) with a total time of 800 ms, see Figure 7.6a.

**Low Signal** The low-signal tone played back when intensity quotients ranged from 1 to 12.9. The tone consists of two consecutive triangle wave pulses with frequencies of 220 and 260 Hz. The amplitude envelope consisted of a 30 ms attack phase, 20 ms sustain phase, and 300 ms release phase.

---

\(^{1}\)Eighteenth-century composer Christian Schubart described the major G chord as evoking “everything rustic, idyllic and lyrical, every calm and satisfied passion, every tender gratitude for true friendship and faithful love - in a word, every gentle and peaceful emotion of the heart is correctly expressed by this key” [41]
7.2. Methods

Figure 7.6: Profile of the auditory display tones, where frequency of individual oscillators is shown on the vertical axis and playback time from 0 to 1000 ms on the horizontal axis. Amplitude envelope is depicted using a gradient from white to dark gray. The zero signal tone employs sine oscillators, and the remaining tones employ triangle oscillators. a) Zero signal, b) low signal, c) high signal, d) error signal.
The 260 Hz pulse played back 233 ms after the start of the 200 Hz pulse. The resulting tone is more intense than the zero-signal tone, as the attack time was shorter, there were more pulses played back, and the triangle wave contained a higher number of upper-level harmonics (resulting in increased ‘brightness’) than the sine pulses of the zero-signal tone; see Fig. 7.6b.

**High Signal** The high-signal tone played back when intensity quotients ranged from 13 to 50. The tone consisted of four sequential triangle waves with frequencies of 349, 440, 523, and 698 Hz. The amplitude envelope of each triangle wave pulse consisted of a 30 ms attack phase, 20 ms sustain phase, and 300 ms release phase. The sequential triangle waves played back with inter-onset intervals of 70 ms. The resulting tone was even more intense than the low-signal tone, indicating an increased urgency due to a higher number of upper-level harmonics and shorter times between sequential pulses; see Fig. 7.6c.

**Error** Finally, the error tone consisted of two simultaneous triangle wave pulses with frequencies of 500 and 515 Hz played back three times with a delay of 100 ms. The amplitude envelope of each pulse consisted of a 20 ms attack phase, 50 ms sustain phase, and 20 ms release phase; see Fig. 7.6d.

Thus, an auditory display synthesizer was created to play back three intensity tones and one error tone. The intensity tones from zero to low to high featured increasing frequency, number of pulses, and onset speed. The four tones were synthesized and played back using the Pure Data [120] sound synthesis environment in real time. Intensity levels were sent at an interval of 800 ms from the fluorescence measurement system to a computer hosting the sound synthesis environment. Playback used a pair of standard multimedia loudspeakers connected to the synthesis computer located approximately 1 m in front of the user.

### 7.2.3 Method Evaluation

#### Participants

Participants (n = 20; male = 10 and female = 10) ranged from ages 23 to 56 (median 25) years old, and none self-reported hearing or vision impairment. The first set of 10 participants were included in the function response test, and the second set of 10 participants were included in the memory test. Self-responded proficiency in music of the first group of participants included 1 professional, 4 amateurs, and 5 with no musical experience. In the second
group, this included 1 professional, 5 amateurs, and 4 with no musical experience. The majority of the participants had backgrounds in engineering or natural sciences but without any professional experience with optical measurements, auditory display, or the project’s background.

**Experimental Setup**

The laboratory experiment setup was designed to assess the ability of participants to distinguish fluorescence readings between three intensity levels and an error signal using the auditory display. This was produced to mimic the ideal situation in the OR, where the surgeon should be informed of intensity levels and error states without an assistant. The experiment consisted of a preliminary intuition test which asked the participants to guess which auditory display tone corresponds to which fluorescence measurement function, and a subsequent function response test in which the participants were asked to measure the fluorescence of a series of phantom wells using the HHF probe and state which function was played back by the auditory display.

**Intuition Test**

For the intuition test, each participant assigned the 4 played tones to functions, including zero signal, low signal, high signal, and error, resulting in 4 data points per participant and 40 data points in total for all participants. To avoid biasing, participants were provided with no information about the application and principles of the measurements. The four tones were played back in a fixed, randomized order used for all participants. First, the tones were played back with each tone repeated three times. Then, the tones were played back again, this time only once. Thereafter, the tones were played back once again, and participants were asked which tones they would assign to each function (zero, low, high or error signal). A final round of playback was undertaken so that participants could change their answer if desired.

**Function Response Test**

During the function response test, participants navigated the HHF probe across a tray of phantom wells (Fig. 7.3) and verbally stated the function perceived for each well. As a training task, participants used a tray orientation that differed from that used later in the test. During training, participants could ask for help or start or stop navigation at will. To complete the training, participants moved the HHF probe over the wells and stated the
perceived tone function after hearing two consecutive, equal readings. The training was performed on 32 training wells, after which all participants confirmed having successfully become familiar with the tones. No response time or accuracy data were recorded during the training phase.

After the training phase, including a 1-min pause, the test procedure was performed for each participant using a 96-well tray (8 rows, 12 columns); see Fig. 7.4. During the test procedure, navigation using the tray was divided into 3 sets of 4 columns each, with a 1-min pause between sets. The actual value and the verbally stated value for each well and each participant were recorded on video.

The tray was filled with phantoms evenly prepared in clusters of 4 values which were arranged across the tray. The sequence of the 24 permutations of the 4 played back tones was randomized [60], and the same sequence was used for all participants; see Fig. 7.4. The phantoms were mixed with an even distribution throughout the well to test all the sequences of tones hearing between zero, low, high, and error signal; see Fig. 7.4. Each well was visually indistinguishable.

To perform the function response task, participants navigated through the wells by moving the tray so that one well was situated directly beneath the probe. The navigation sequence was such that participants measured all wells in one column and then moved on to the next column. After the fluorescence was measured by the HHF probe, tone was played back through the loudspeakers. As in the training phase, participants were instructed to only state the perceived function after being confident of its stability by listening for two consecutive, equal tones using the auditory display. After stating the perceived function (“zero,” “low,” “high,” or “error”) the participant moved the tray so that the next probe measurement could be taken. This process was repeated until all wells in the tray had been measured and the perceived functions stated. The subjects were not provided with any visual clues of the measured signal while evaluating the tones.

**Memory Test**

The intuition test was performed at the first step. The participants were then instructed on the intended function and trained by playing the tones in a random order [60], after which the participants gave their response. This was repeated for approximately five rounds of playback. The memory was tested on days 1-3 and days 7-12 depending on the participant availability. For the memory test, the tones were played in a random order once and afterward one by one in the same order when the participant responded.
### 7.2.4 Data Analysis and Statistics

For each phantom well, the played sound value was compared to the participant’s response. Using power sample size calculation, a minimum of 55 samples were needed to achieve 98% accuracy with a power of 0.95. The number of measurements for each participant was approximately twice as much as this value. In total, 960 data points were recorded. The total number of played tones for each function were 255 zero signals, 263 low signals, 269 high signals, and 173 error signals. No data were excluded from the analysis.

Accuracy was calculated as the ratio of the total number of correctly identified tones to the total number of played tones for each tone and each participant. The latency of the response was calculated by the number of pulses (single, played back tone) needed until the participant uttered the response. The number of pulses was recorded, including the minimum two initial pulses, that the participants needed before uttering a response. Pulses uttered during playback of a tone were recorded as having the previous number of pulses. For instance, if a participant responded while the fourth pulse played back, this was recorded as a latency of 3 pulses.

The statistical tests were performed in MATLAB® R2015a (MathWorks™, Inc.). The null hypothesis was that the played sounds were not distinguishable and that there was no correlation between the played and perceived sounds. As some of the datasets were not normally distributed, the Mann-Whitney test was used for assessing statistically significant difference, where $p < 0.05$ was considered to show a statistical significance. Linear correlation was used for assessing the goodness of fit ($R^2$) between each of the two datasets. Boxplots were used to represent the data sets where the mid-line in the box was the median and the box was set to 25% to 75% quartiles.
Chapter 7. Fluorescence-Guided Open Brain Tumor Surgery

### 7.3 Results

#### 7.3.1 Intuition Test

Twenty-one of the 40 reported tone intuition assignments corresponded to the intended function, resulting in a true-to-total ratio of 52%; see Table 7.2. Tone two was the most intuitive, as all participants correctly assigned this to the high-signal function. Tones zero and one were the least intuitive; 50% assigned the zero signal tone to low-signal function, and 60% assigned the low-signal sound to the zero-signal function. For the error tone, 60% correctly assigned the correct signal, whereas 20% assigned this to the zero-signal function and 20% to the low signal function.

<table>
<thead>
<tr>
<th>Perceived</th>
<th>Played</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero</td>
<td>3 6 0 2</td>
</tr>
<tr>
<td>One</td>
<td>5 2 0 2</td>
</tr>
<tr>
<td>Two</td>
<td>0 0 10 0</td>
</tr>
<tr>
<td>Error</td>
<td>2 2 0 6</td>
</tr>
</tbody>
</table>

**Table 7.2:** Confusion matrix for intuition test results showing played versus perceived functions of the auditory display tones

<table>
<thead>
<tr>
<th>Perceived</th>
<th>Played</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero</td>
<td>247 2 0 0</td>
</tr>
<tr>
<td>One</td>
<td>7 258 1 1</td>
</tr>
<tr>
<td>Two</td>
<td>0 3 267 1</td>
</tr>
<tr>
<td>Error</td>
<td>1 0 1 171</td>
</tr>
</tbody>
</table>

**Table 7.3:** Confusion matrix for function response accuracy showing played versus perceived functions of the auditory display during laboratory evaluation measurement.
7.3. Results

Figure 7.7: Median of latency versus accuracy for each participant. No correlation was found between the two parameters. The musical proficiency of the participants is notated beside each data point, where 0 is no musical training, 1 is amateur training, and 2 is professional training.

7.3.2 Function Response Test

Response Accuracy

The training period for the 10 participants lasted an average of 80.44 (±29.44) s. After training, 943/960 (98%) of the participants’ responses matched the played function. Recorded values versus verbally given responses are shown as a confusion matrix in Table 7.3. The highest confusion was exhibited for the zero-signal tone: It was perceived in 3% of the cases to be the low-signal tone. The high-signal tone was the most confidently perceived, as more than 99% of responses were correct. There was no statistically significant difference between the played tone and the responses for any of the participants when each played tone was compared to each verbal response for each participant (p-value for each participant > 0.8). There was a low and negative correlation between the median accuracy and the musical proficiency for each participant ($R^2 = 0.3$), but no correlation was found between the accuracy in the initial intuition test before the training and the accuracy after training ($R^2 < 0.1$).
Response Latency

The response latency (time needed to respond verbally) was a minimum of 2 pulses (1.6 s), and at most 9 pulses (7.2 s) with a median of 2 (average = 2.6) pulses across all participants. The response latency for each participant is plotted in Fig. 7.8. When the latency for each tone was separately analyzed for each participant, only in participant no. 6 was the median different for separate tones; therefore, the latency showed a dependence on the individual rather than on the tone. The correlation between the median latency and accuracy ($R^2 < 0.1$) for each participant was negligible (Fig. 7.7). The musical proficiency (0, 1, 2) did not show any correlation with the median response latency for each participant ($R^2 = 0$).

Memory Response

The initial intuition test for this population was 60% accurate. After becoming acquainted with the tone functions, the responses during the training were 100% accurate both on days 2 to 3 and days 7 to 12.
7.4 Discussion

The goal of this study was to investigate the practicality of using auditory displays for the communication of PpIX measurement results based on those used during neurosurgery. The experimental study was designed to determine how well the intensity of a fluorescence measurement using an HHF probe could be recognized by listening to tones played back in a laboratory environment. The auditory display for fluorescence intensity was developed to provide the surgeon with four discrete tones to help identify the status of the fluorescence measurement without having to view a computer monitor or rely on the support of an assistant to verbally relay intensity values. The principles can be applied to any other PpIX fluorescence spectral analysis algorithm; however, the limits of zero, low, and high signals would depend on the device and the analysis algorithm [2, 45, 88, 143].

7.4.1 Function Response

The participants’ intuitively assigned tone functions were evaluated, as well as how quickly and accurately participants could respond to the played tones while measuring a tray of fluorescence phantoms. The results of the evaluation show that in almost all cases, participants correctly identified each of the four played intensity tones after a median of 2 pulses (1.6 s), where two pulses was the minimum amount of time to correctly provide an intensity measurement. The intuition test showed that even before using the system or receiving training, participants were able to correctly assign the synthesized sounds to the intended functions in more than half of all cases. After training, the response accuracy increased to 98% for the function test, and the memory test was 100% accurate for all participants after periods of both 2-3 days and 7-12. Thus, the developed auditory display is easy to learn and remember, quick to use, and highly accurate. Even in its current form, the auditory display could provide an immediate benefit to surgeons by delivering intensity values without the need for a screen or reliance on a surgical assistant to relay values.

The participants were not provided with any information on the principles of fluorescence or the intended purpose of the measurements to reduce the effect of being assisted by vision. The unintentional visual assistance which might have been induced in some cases at the phantoms with low and high PpIX concentration was considered negligible as the participants could not interpret the colors and the colors were blurry without any optical filter.
Measurement accuracy had a median of 66% (range 35-95%) weakening any chances of visual assistance. This measurement error which was caused due to the wall of the phantom wells does not occur in the actual situation during operation.

### 7.4.2 Auditory Display Design

The auditory display should be designed to be easy to learn, and tones should be distinguishable from one another, so that the surgeon is able to understand the meaning of the display in a few seconds. The auditory display should be fast enough to transmit the desired intensity to the surgeon within the duty cycle of the system, i.e., the playback of the entire tone should occur between successive measurements. The results of the evaluation show that on average, participants could recognize the sound and verbally respond with the correct value directly within the duty cycle of the system even after undergoing only a brief training period.

The auditory display should be easily heard in the OR alongside other sounds [42], being sufficiently distinguishable from but not interfering with existing sounds, such as those form suction devices, anesthesia equipment, or ICU sounds. The auditory display should not sound similar to an alarm, as the transmission of intensity levels using auditory display is not an alarm, and surgeons could become annoyed when presented with sounds that are perceived to be unnecessarily urgent [157]. Indeed, unnecessarily urgent sounds can become quickly fatiguing [115], and common auditory signals in clinical settings have been shown to convey an unintended, inappropriate level of urgency [108]. Although IEC 60601-1-8 is an international standard that addresses alarms in OR and categorizes these into low, medium, and high priorities, implementation of the standard is controversial [38, 43, 130]. Adherence to the standard is not compulsory although the medical alarm manufacturers are likely to comply [43]. The current work, however, is distinguished as an auditory display to be used actively by the surgeon, as opposed to an alarm, which merely notifies clinical staff at some point when a threshold has been reached. Thus, the auditory display for HHF probe measurement does not fall under the auspices of IEC 60601-1-8. Currently, there is no accepted standard for the design of non-alarm auditory displays in the OR, likely due to the paucity of investigations into the field of auditory display in interventions [16].

Although the preliminary results of this study are very promising, there are open questions regarding the implementation of such an auditory display system in the OR. First, although this specific application categorizes intensity values into four different levels (zero signal, low signal, high signal, and
error), auditory display could also be enhanced to accommodate either fewer or more intensity levels. This number could dynamically change depending on the current clinical requirements or familiarity of the clinician with the auditory display. For instances in which the clinical scenario is straightforward and requires only a binary decision support, the auditory display could, for example, be reduced to provide simply zero- and high-level tones. However, when the scenario is more complex, additional levels could be provided. The most relevant application for continuous sound is for fluorescence measurement during stereotactic biopsy [66, 103]. In this application, it is desired to map the tumor, i.e., to provide information on the availability and intensity of the fluorescence along the stereotactic insertion path where the sites with the highest fluorescence signal are selected for biopsy. This could provide a stronger measure of confidence for these areas for which the clinician must be sure of eligibility for removal. A second option would be to encode the intensity using a hybrid approach that combines continuous mapping with the employed level-based tones. Thus, for instance, 1, 2, or 3 primary tones (such as those described in this work) could be augmented with secondary auditory parameter mapping such as the intensity of vibrato (slight frequency modulation) or tremolo (slight volume modulation). This could strike a balance between a quick classification of fluorescence intensities as well as provide clinicians with a more nuanced way of detecting critical areas at tumor boundaries. In any case, the clinical scenario and clinician familiarity with the auditory display should determine the appropriate method to be employed. Issues of certification must be addressed as well, as a binary classification would place a higher burden on the device manufacturer, whereas a continuous or multi-level approach would place a higher burden on the clinician to discriminate between tumor and healthy tissue.

In terms of auditory display design, the most frequently incorrectly perceived sound in the study (zero signal) could be redesigned to be more unique, thereby possibly increasing recognition rates by participants. Second, the latency of the fluorescence measurement might be accelerated so that the surgeon could receive intensity information even faster. Third, although the auditory display was specifically developed to produce tones that should not interfere with common existing OR sounds, the integration of the auditory display in the OR must be further customized to the specific environment in which it is to be played. This should minimize interference with sounds from other devices. Integration must also take playback mechanisms into account; for instance, a small loudspeaker placed locally near the surgeon could play back the tones rather than a speaker placed somewhere further away in the room.
7.4.3 Future directions

Future development should take the aforementioned factors into account and evaluate the auditory display in an environment that is even closer to that of the actual clinical scenario or in the OR. In addition, time, accuracy, and workload aspects should be compared to those cases in which an assistant relays intensity values to the surgeon to better discern the benefit of such an auditory display. Methods such as augmented reality and surgical microscope image injection will surely play a major role in the operating rooms of the future. However, this is often accompanied with negative impacts on clinician attention, for instance, significantly reducing foreign body recognition in endoscopic viewing supported by augmented reality [40].

The development of a hybrid display system could be beneficial, for instance, by transmitting the intensity of the quantified fluorescence with both auditory and visual means. Thus, future evaluations should determine in which cases each method of feedback is most suitable. To the authors knowledge, there have been no such comprehensive evaluations in the clinical field. An ideal subsequent solution could be a harmonization in a hybrid system that provides both immediate, relevant intensity information through auditory display and more complex details using visual display, either through a screen or using augmented or virtual reality concepts.

7.5 Conclusion

The experiments show that the employed auditory display is fairly intuitive, easy to learn, accurate, and fast in providing the user with the measurement of the current intensity or error signal for fluorescence-guided resection. Auditory display is a nascent field that could bring real benefit to surgeons using FGR, reducing reliance on a computer screen or surgical assistant and allowing focus to be retained on the surgical site. Future work should refine auditory methods and evaluate the concept in a more realistic clinical environment as well as investigate possible combinations with visual methods such as augmented and mixed reality.
Chapter 8

Sterile Interaction Using Eye Tracking


About this chapter

The growing number of technical systems in the operating room has increased attention on developing touchless interaction methods for sterile conditions. However, touchless interaction paradigms lack the tactile feedback found in common input devices such as mice and keyboards. This chapter introduces auditory display for a touchless eye-tracking interaction system for completing typical scrub-nurse tasks in the operating room. Due to the absence of tactile feedback for eye-tracking and other touchless interaction methods, auditory display is shown to be a useful and necessary addition to new interaction concepts for the sterile operating room, reducing reaction times while improving subjective measures, including usefulness, user satisfaction, and cognitive workload.
Chapter 8. Sterile Interaction Using Eye Tracking

8.1 Eye Tracking for Sterile Interaction

With the increasing number of computer systems available in today’s operating rooms, both clinicians and assistants are exposed to a plethora of interaction possibilities in both preinterventional and intraoperative phases. Although this increase endows clinicians with expanded and enhanced interaction possibilities, the sterile nature of the operating room often limits or prevents the use of traditional computer input devices such as mouse and keyboard, as these can be quickly contaminated during an operation [128]. Touchscreens, joysticks, and buttons are typically covered in a plastic sheath, which become soiled during an operation and must often be changed, leading to costly delays [140]. In many complex cases, clinicians must exit the sterile environment to operate a workstation, which requires repeated scrubbing into the operating room.

Thus, a current trend in operating room interaction research is to investigate new paradigms for touchless interaction. Current solutions for sterile interaction in the operating room include foot pedals [49, 70] and gesture-controlled interaction systems using standard motion-detection devices such as the Leap Motion controller [10, 15] or the Microsoft Kinect [89]. For a review of the literature on touchless interaction in the operating room, see Mewes et al. [106]. Gesture-based interfaces promise improved sterile interaction with computer systems in the operating room, although their use has not yet been established in clinical routines. Although these methods free the clinician or assistant from having to touch input devices, sterile interaction systems based on gestures exhibit the drawback that interaction with the device is based on hand or body movement. By doing so, a user is required to interrupt the clinical routine to interact with such a system. To this end, the use of eye-tracking glasses as a sterile means of interaction have been investigated to eliminate the need for complex hand or body gestures. This has already been successfully demonstrated, for instance, in laparoscopic surgery for training [34] or to position the camera [100]. Furthermore, use cases for sterile eye tracking have been proposed for scrub nurses [93].

Although eye-tracking interaction systems show great potential for flexible employment in sterile environments, they, along with other touchless interaction concepts, only provide primary visual feedback, in that the user can only see the result of an initiated action on a screen. Unlike traditional input devices such as keyboard, mice, and joysticks, touchless interaction concepts, thus, fail to provide the user with a secondary, tactile means of feedback to inform the user that an action has been performed. This may impede the adoption of touchless interaction concepts in the operating room, as the availability of real-time feedback for interaction has been shown to reduce mental load and improve efficiency and user satisfaction [82].
To overcome the lack of a secondary feedback in touchless interaction concepts, the use of auditory display has been proposed to provide the user with this missing secondary feedback. Compared to visual-only feedback, the addition of auditory display reduces selection time and improves selection accuracy in hand gesture-based circular menus [113] using the Leap Motion Controller as input device.

This work investigates the benefits of both eye-tracking as an input concept for sterile operating room interaction and the effect of the inclusion of auditory display as a secondary feedback mechanism. We propose a touchless, gesture-free, sterile interaction method using eye-tracking to help surgical assistants complete a series of 3 typical OR uses cases, each with 2 tasks.
8.2 Methods

Essential to the interaction with the eye-tracking system are so-called gaze gestures, which describe how the user views objects in the scene. In the developed eye-tracking system, QR (quick response) codes were printed and placed within the demonstration operating room to permit study participants to view locations and initiate actions with the system. Two kinds of gaze gestures were created to permit interaction: simple gaze gestures and repeated gaze gestures.

Simple gaze gestures describe a single trigger when a user views the desired QR code for a defined amount of time. This duration must be long enough to allow the eye-tracking system to recognize the gaze and prevent accidental triggering of actions when the user simply changes view around the demonstrator operating room, but short enough to provide a sufficiently rapid feedback response time so that the user is not frustrated by excessive latency. In the case of the developed sterile interaction system, preliminary design iterations with potential clinical users resulted in a system recognition time of 1 second to strike an optimal balance between robust recognition, minimal accidental gaze gesture occurrence, and short enough interaction time.

Repeating gaze gestures are those for which a user initiates a continuous action with the eye-tracking system that extends over a variable period of time. Repeating gaze gestures are executed by viewing a QR code and holding the gaze until interaction is completed. This method is used in this evaluation, for instance, to dim operating room lights. During light dimming, the repeated gaze gesture continuously controls the brightness of the light bulb in stages between complete darkness and maximum brightness.

8.2.1 Eye-tracking System Design

The developed eye-tracking system employs a set of SMI v2.0 eye-tracking glasses, which consist of an outward-facing camera with a resolution of 1280 \times 960 pixels combined with inward-facing infrared cameras which detect the point of regard (PoR) of the user in relation to the forward-facing captured scene. The glasses weight 68 grams and are connected by a USB cable to a host computer.
8.2. Methods

The PoR as well as the images acquired by the forward-facing visual camera are retrieved by proprietary software which is delivered with the eye-tracking glasses. This custom software extracts the QR codes from the visual image, determines if the PoR is located on a QR code and sends a network message to client software when a gaze gesture is detected. The client software handles the interaction with the devices in the OR, including room lights, telephone, and surgical instrument management system.

8.2.2 Auditory Display as Secondary Feedback

The auditory display system for eye-tracking interaction provides a secondary feedback to the system, which is missing due to the touchless nature of the interaction. This is contrasted to interaction with physical devices such as mice, keyboards, joysticks, and buttons, all of which are commonly used in the operating room and provide a secondary, tactile feedback, produced when the user actuates the device. This tactile feedback is useful to know whether or not an action was successfully performed. Touchless interaction concepts, however, most commonly rely on hand gestures or eye gaze, neither of which provide a secondary feedback mechanism necessary to ensure the user that a gesture was recognized. To augment a touchless interface with secondary feedback, auditory display has been shown to increase efficiency as well as satisfaction [16]. In addition to informing the user that a gesture or gaze was recognized, auditory display can also inform the user as to which specific gesture or gaze was recognized, so that undesired input can be corrected, thereby easing training with the system.
To support secondary feedback for use cases, an auditory display environment was created to support both the simple gaze gesture use cases 1 to 4 (begin and end video call, marking instrument, and generating instrument report) as well as the repeating gaze gesture in use cases 5 and 6, dimming the operating room light to full brightness or complete darkness. For all auditory display, the PureData \cite{120} real-time audio programming environment was used.

8.2.3 Use Cases for Sterile Interaction

Based on Glaser et al. \cite{93}, who proposed suitable use cases for potential interaction with eye-tracking in a sterile environment, we chose 3 typical use cases found in the OR that are often performed by surgical assistants. These use cases were derived from the situation of a surgical assistant or scrub nurse standing next to the operating table and in front of the instrument table. For each of the use cases, the participant could complete 2 tasks by initiating an action by gazing at a QR code located in the test laboratory.

Use Case 1: Initiating and Ending a Video Call The use cases required the scrub nurse to (task 1) initiate a video or telephone call for the clinician and (task 2) end the call thereafter.

Use Case 2: Marking an Instrument as Defective and Generating an Instrument Report After being handed an instrument, the scrub nurse must mark (task 3) the instrument as defective so that it can be replaced or repaired during sterilization after the operation. The scrub nurse prepares an electronic report of the defective instrument received. Here (task 4), the scrub nurse sends the report to the clinical database.

Use Case 3: Controlling Operating Room Light Brightness In these cases, the surgical assistant controls the brightness of an operating room lamp by (task 5) dimming to complete brightness and (task 6) dimming to complete darkness.

8.2.4 Earcons for Simple Gaze Gestures

For simple gaze gesture tasks 1 to 4, we implemented a set of so-called earcons for informing the user when a gesture was recognized by the eye-tracking system, and, specifically, which gesture was recognized and executed. Earcons are short, abstract, synthetic musical audio segments used
to deliver messages about events in a system [72]. These short segments, usually single pitches or rhythmic sequences of pitches, may be grouped into larger units called families of related motives to identify related messages. Audio parameters such as timbre, register, pitch, rhythm, duration, and tempo are useful for differentiating earcons and grouping them into families. During initial design iterations with potential users, various earcon variations were generated and evaluated using informal think-aloud protocols, resulting in the single set of 5 earcons described here and employed in the rest of the evaluation. The fixed set of earcons was used as a basis to determine the overall feasibility of the approach. The earcons employed melodies and chords found in Figure 8.3 using a software digital synthesizer with a marimba tone. Each earcon lasted less than 1 second.

To support feedback of gaze recognition, earcons were generated for the following interactions:

- Initiation of simple gaze gesture
- Generate instrument report, simple gaze recognized
- Video call begin, simple gaze recognized
- Video call end, simple gaze recognized
- Instrument defective, simple gaze recognized
8.2.5 Parameter-Mapping Auditory Display for Repeated Gaze Gestures

In the case of auditory display for dimming the operating room lamp (tasks 5 and 6), a parameter-mapping auditory display mechanism was implemented. In parameter-mapping auditory display, the underlying data are used to ‘play’ a real-time software instrument according to those changes. Because audio has various parameters that may be altered (such as frequency, intensity, and timbre), continuous parameter mapping is suitable for displaying multivariate data. This technique makes the listener an active participant in the listening process by browsing the data set using the auditory display or by interactively changing the mappings that relate data to audio [72]. The listener can navigate through a set of data to perform a task. Thus, this method is useful for smoothly representing continuous changes in underlying data. In the case of light dimming, the auditory display conveys the current light brightness using a parameter-mapping mechanism. The brightness of the bulb is mapped to the pitch of a sine tone generator quantized to notes of a major C-scale from MIDI note 60 (C, 261 Hz) to note 72 (C, 523 Hz). During brightness changes, the moving tone alternates with a steady sine tone of 523
Hz at a frequency of 2 Hz to provide the user with the ability to compare both tones. By doing so, the user receives feedback on the current brightness relative to maximum brightness. When reaching full brightness (bulb on) or complete darkness (bulb off), a major chord is played back.

### 8.2.6 Experiment Design

We conducted a user study employing the prototypical eye-tracking system to evaluate the effect of auditory display as a substitute for the loss of tactile feedback during sterile interaction tasks with respect to response time, subjective workload, and system acceptance.

#### Physical Setup

A laboratory operating room was outfitted with the eye-tracking system. The operating room consisted of an operating table, instrument table, and three main computer displays. The eye-tracking system was installed on a computer which sat out of sight of the participant. The participant sat on a stool next to the operating table and in front of the instrument table, facing three main computer displays. A speaker to play the generated auditory display synthesis output was placed opposite the participant near the operating table. As a means of visual feedback, full-screen icons that were shown after completing each task were displayed on the central computer monitor in the laboratory operating room.

#### Demographic Composition

Twenty-six (26) participants completed the laboratory evaluation. The 18 male and 8 female participants had an average age of 29 (ranging from 24 to 43), and 9 wore glasses. One participant reported a slight hearing disability. Three participants came from medical backgrounds with experience in an operating room and the remainder were scientific researchers. Half of the participants were assigned to group AV (first performing the tasks with audiovisual feedback, followed by visual-only feedback) and the other half to group VA (first performing the tasks with visual-only feedback, followed by audiovisual feedback).
Procedure

The experiment consisted of a set of 6 tasks which represented use cases found in a typical ear-nose-throat intervention, including initiating a video call, hanging up the video call, generating an instrument report, labeling an instrument as defective, turning a light on, and turning a light off. Each participant repeated the entire set of tasks once using audiovisual feedback and once using visual-only feedback. Before beginning the experiment, participants signed an informed consent agreement, the experimenter explained the usage of the eye-tracking system to the participant, and a calibration of the eye-tracking glasses was completed. For both audiovisual and visual-only feedback, each participant completed a training session of 12 tasks. Thereafter, the participants completed the set of 6 tasks 3 times each in a random sequence, thus resulting in 24 training and 36 test tasks completed per participant. During all tasks, an audio recording of a typical ENT surgery was played in the background.

For each individual task, participants were instructed by the experimenter, who played the role of the lead surgeon during an ENT intervention, to execute the desired task. After instruction, the participant pressed a timer button located on the instrument table, thereby recording the start time of the task. After performing the desired task, the participant again pressed the timer button to indicate task completion.

Following the entire sequence of tasks using either audiovisual or visual-only feedback, the participant completed a questionnaire consisting of demographic questions and three sets of evaluation questions. Van der Laan technology acceptance scale [94] was used to judge the usefulness of and satisfaction with using the novel eye-tracking system. The questionnaire provides 9 pairs of adjectives, such as “effective/superfluous” or “assisting/worthless” to generate composite ratings for usefulness and satisfaction. The widely-used NASA Task Load Index (TLX) scale [31] was used to measure the workload experienced by the participants during task completion. We employed the Raw TLX scale, as the weighted scale is more time intensive and has not been conclusively shown to provide additional benefits [69]. Finally, four additional questions were asked concerning task execution confidence, ease of use, necessary time to complete the task, and helpfulness of the feedback method.
8.2.7 Statistical Analysis

Reaction time was recorded for each executed task, determined by the elapsed time between the point the system triggered the designated action and the participant engaged the timer button at end of the task. We excluded reaction times under 0.25 seconds because reaction time would have been less than a physiologically reasonable reaction time [78]. In such cases, users preemptively pressed the task-complete button before the eye-tracking system could successfully complete the task and provide an auditory or visual feedback.

Reaction times for each task (dependent variable) were analyzed as mixed Analysis of Variance (ANOVA) of feedback (audiovisual, visual only) as within subjects factor (conducted by all participants) and group (AV, VA) as between subjects factor (either starting with audiovisual or with visual feedback and conducting the second condition thereafter), see Table 8.1. This was done to determine the presence of training effects on reaction time by receiving additional training during the first completed condition. The questionnaires for user feedback were analyzed by repeated measures ANOVA for the condition of feedback (audiovisual, visual only), irrespective of the group, since no training effects were expected for subjective feedback. The level of significance, i.e., the statistical difference of the means, was indicated by $p < 0.05$. For details on the ANOVA method used, see [47].

8.3 Results

Reaction Time  The average reaction times for audiovisual and visual-only feedback are shown in Figure 8.5. Average reaction times were significantly faster when using audio for all use cases, except for dimming the light off, where a level of significance was not reached. A slight training effect occurred for the light-off task, in which the VA group (first using visual-only feedback and then audiovisual) improved reaction time for the second round more than the AV group (first using audiovisual feedback and then visual-only feedback). In addition, the task of placing a video call also showed significant group interaction effects; the AV group maintained their reaction time from the first round (A) to the second (V), and the VA group improved their reaction time from the first round (V) to the second round (A).
## Table 8.1: Summary of statistical analysis of repeated ANOVA for reaction times (above) and questionnaires (below)

### Table 8.1: Summary of statistical analysis of repeated ANOVA

<table>
<thead>
<tr>
<th>Task</th>
<th>Condition</th>
<th>dF</th>
<th>F</th>
<th>p</th>
<th>sig</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call</td>
<td>time</td>
<td>1, 24</td>
<td>5.14</td>
<td>0.03</td>
<td>*</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>6.63</td>
<td>0.017</td>
<td>*</td>
<td>0.22</td>
</tr>
<tr>
<td>Hangup</td>
<td>time</td>
<td>1, 24</td>
<td>35.35</td>
<td>0.00</td>
<td>**</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>3.16</td>
<td>0.09</td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>Generate Report</td>
<td>time</td>
<td>1, 24</td>
<td>6.10</td>
<td>0.02</td>
<td>*</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>0.01</td>
<td>0.94</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Mark Defective</td>
<td>time</td>
<td>1, 24</td>
<td>73.53</td>
<td>0.00</td>
<td>**</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>2.20</td>
<td>0.15</td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td>Light On</td>
<td>time</td>
<td>1, 24</td>
<td>32.66</td>
<td>0.00</td>
<td>**</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>1.11</td>
<td>0.30</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>Light Off</td>
<td>time</td>
<td>1, 24</td>
<td>1.13</td>
<td>0.30</td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>time*group</td>
<td>1, 24</td>
<td>6.03</td>
<td>0.02</td>
<td>*</td>
<td>0.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>dF</th>
<th>F</th>
<th>p</th>
<th>sig</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA Raw TLX (overall)</td>
<td>1, 25</td>
<td>8.77</td>
<td>0.01</td>
<td>**</td>
<td>0.26</td>
</tr>
<tr>
<td>van der Laan (overall)</td>
<td>1, 25</td>
<td>10.21</td>
<td>0.00</td>
<td>**</td>
<td>0.29</td>
</tr>
<tr>
<td>Usefulness</td>
<td>1,25</td>
<td>7.57</td>
<td>0.01</td>
<td>**</td>
<td>0.23</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>1,25</td>
<td>8.9</td>
<td>0.01</td>
<td>**</td>
<td>0.26</td>
</tr>
<tr>
<td>Agreement (overall)</td>
<td>1</td>
<td>24.89</td>
<td>0.00</td>
<td>**</td>
<td>0.50</td>
</tr>
<tr>
<td>Confidence</td>
<td>1, 25</td>
<td>29.56</td>
<td>0.00</td>
<td>**</td>
<td>0.54</td>
</tr>
<tr>
<td>Ease</td>
<td>1, 25</td>
<td>3.22</td>
<td>0.09</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>Time</td>
<td>1, 25</td>
<td>5.00</td>
<td>0.03</td>
<td>*</td>
<td>0.17</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>1, 25</td>
<td>40.10</td>
<td>0.00</td>
<td>**</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Chapter 8. Sterile Interaction Using Eye Tracking
8.3. Results

### Workload
The results of the NASA Raw TLX questionnaire are shown in Table 8.2. The mean overall subjective workload for the audiovisual feedback was 32.50, whereas the mean subjective workload for visual-only feedback was 40.23. Interaction with audiovisual feedback showed significantly reduced subjective workload overall compared to interaction without auditory display. In every dimension, auditory display was rated with less cognitive load compared than when using visual-only feedback.

### System Acceptance
The results of the van der Laan system acceptance questionnaire [94] are shown in Table 8.2. Using the Van der Laan scale, both the usefulness as well as satisfaction with the eye-tracking system were rated as significantly higher when using audiovisual feedback as opposed to visual-only feedback.
Chapter 8. Sterile Interaction Using Eye Tracking

Table 8.2: Mean and standard deviation values for questionnaire results over all participants for both audiovisual and visual feedback. Ranges for NASA Raw TLX are 0 (low workload) to 100 (high workload), for van der Laan -2 (fully reject) to +2 (fully accept) and for agreement questions 0 (strongly disagree) to 7 (strongly agree).

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Audiovisual M (SD)</th>
<th>Visual M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASA Raw TLX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>32.50 (13.93)</td>
<td>40.23 (12.94)</td>
</tr>
<tr>
<td>Mental</td>
<td>32.31 (19.91)</td>
<td>40.58 (25.50)</td>
</tr>
<tr>
<td>Physical</td>
<td>23.27 (21.21)</td>
<td>29.62 (21.91)</td>
</tr>
<tr>
<td>Temporal</td>
<td>37.50 (21.65)</td>
<td>44.04 (23.83)</td>
</tr>
<tr>
<td>Performance</td>
<td>25.00 (21.21)</td>
<td>27.88 (12.90)</td>
</tr>
<tr>
<td>Effort</td>
<td>58.85 (26.84)</td>
<td>65.77 (23.05)</td>
</tr>
<tr>
<td>Frustration</td>
<td>18.05 (26.84)</td>
<td>33.46 (21.39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Laan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usefulness</td>
<td>1.26 (0.62)</td>
<td>0.88 (0.61)</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>1.11 (0.66)</td>
<td>0.69 (0.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>5.35 (0.69)</td>
<td>3.96 (1.51)</td>
</tr>
<tr>
<td>Ease</td>
<td>5.23 (0.95)</td>
<td>4.69 (1.23)</td>
</tr>
<tr>
<td>Time</td>
<td>5.00 (1.33)</td>
<td>4.50 (1.42)</td>
</tr>
<tr>
<td>Helpfulness</td>
<td>5.27 (1.08)</td>
<td>2.73 (1.89)</td>
</tr>
</tbody>
</table>

Participant Agreement  The results of questions regarding participant satisfaction with confidence, ease of use, and helpfulness of the display are shown in Table 8.2. Audiovisual feedback was rated as providing significantly higher satisfaction than when using visual feedback for confidence, ease of use, and helpfulness. For satisfaction with the amount of time needed to complete the task, no significant difference was found.
8.4 Discussion

We developed a novel concept to interact with operating room technology that is both sterile and does not require the use of hands or feet for input. Touchless interaction systems are desired in the operating room to increase the possibilities of using certain systems without the need for sterilization of physical input devices. By implementing touchless interaction, surgical assistants or clinicians could save valuable time and costs by reducing the amount of sterilization that is involved with ordinary input devices.

However, because touchless interaction paradigms do not include the secondary haptic feedback generated when interacting with typical operating room input devices such as mice, keyboards, buttons, and joysticks, an auditory display was created to provide feedback to compensate for this loss.

The developed auditory display supports interaction with the eye-tracking system by notifying the user when a gaze target is recognized by the eye-tracking system, by providing feedback regarding which gesture was successfully recognized, as well as by generating continuous feedback regarding repeating gestures such as dimming operating room lights. The auditory display consisted of a set of earcons for simple gaze gestures for tasks including starting and ending a video call, generating an instrument report, and generating the report. Continuous auditory feedback was generated to relay the brightness of an operating room light bulb with notifications when the light was completely off or at full brightness.

The evaluation of the eye-tracking system compared the effects of audiovisual and visual-only feedback on reaction time, subjective workload, usefulness, and satisfaction for 6 common surgical assistant tasks.

For each of the subjective dependent measures, audiovisual feedback was shown to be superior to visual-only feedback. Average reaction time was significantly reduced when using the audiovisual feedback for all tasks except for turning the operating light off, for which reaction time was reduced but a statistical level of significance could not be reached. The largest average time saving using audiovisual feedback was found during the light-on task, perhaps because the light dimming function gives no feedback when full brightness is achieved. Thus, when changing lights to full brightness, participants may have achieved full brightness without having realized this. Completing such a task without receiving feedback could be especially frustrating, as noted in comments by some of the participants.
In addition to improving reaction time, audiovisual feedback was rated significantly higher for system acceptance, significantly lower for workload, and significantly higher for participant satisfaction for confidence, ease of use, and helpfulness of the feedback methods.

Implementing audiovisual display in the operating room is not without challenges. By design, operating rooms host a range of sound sources, such as speech and instrument noises, as well as functional warning and status monitoring sounds emanating, for example, from anesthesiology equipment. Previous studies [22, 36, 67, 90, 157] which evaluate auditory display in the operating room have noted the need for auditory display design to take existing operating room sounds into account. However, despite the operating environment often being noisy, according to Katz et al. [84], “there is little evidence to demonstrate a direct association between excessive operating room noise and poor surgical outcomes.” Additionally, Moorthy et al. [109] report that surgeons can successfully disregard extraneous noise in the operating room. However, to minimize extraneous noise in the operating room generated, for instance, by the auditory feedback described in this work, a variety of playback methods can transmit the desired auditory feedback while permitting communication between surgical team members. These could include bone-conducting headphones, open-back headphones, small speakers located near the scrub nurse, or even parabolic speakers which focus sound output directly to the target nurse.

Further investigations must evaluate which sound output method provides the best usability while generating the least amount of noise and distraction to other team members. In addition, customized auditory display could play a role in a networked operating room (such as OR.NET, see Rokstroh et al. [124]), in which each member of a complex, interwoven intervention team receives personalized auditory feedback based on individualized auditory display preference and current task. In addition, further evaluations should determine the extent to which variations of the auditory display (including, for instance, aforementioned personalized earcons or voice samples) affect performance and subjective workload. For clinical scenarios in which a high number of tasks must be completed by the nurse, a subsequent evaluation could similarly determine the effect of auditory display on performance and subjective workload. A further factor that must be considered in future auditory display designs is the total number of sounds to be incorporated: because earcons are abstract and need to be trained, a balance must be struck between the number of earcons and the ability of the nurse to train and remember them successfully during the clinical routine. If the number of functions to be employed using auditory display is large, grouped or
hierarchical earcons [21] could be developed to accelerate training. Doing so, similar clinical task functions could be grouped to correspond to similar earcons, thus hypothetically increasing extendibility and reducing initial training time compared to many dissimilar earcons.

8.5 Conclusion

As the number of technological possibilities within the operating room continues to expand, touchless interaction will be a field of focus as a sterile input method. This evaluation of prototypical eye-tracking for touchless interaction in the operating room demonstrates that such a system is best used in combination with an audiovisual display rather than a purely visual display. Auditory display has demonstrated to improve quantitative performance measures and exhibits lower workload and higher satisfaction and acceptance than visual-only feedback. Touchless interaction for gesture recognition is not yet an established method within the operating room, and its introduction will be burdened if there is no secondary feedback mechanism present. The use of auditory display or audiovisual display will help accelerate implementation and acceptance of novel touchless interaction in the sterile operating room.
Chapter 9

Freehand Gesture Interaction


About this chapter

In addition to the novel eye-tracking system described in the previous chapter, free-hand gesture recognition technologies also allow touchless interaction with a range of applications within medicine, such as scrolling image slices or activating system commands. Similar to interaction using eye-tracking, touchless freehand using, for instance, a Leap motion controller, usually only provide feedback on a screen. This work explores auditory display with auditory icons and continuous, model-based sonification to improve free-hand gestures using a Leap motion controller. Three concepts were generated and evaluated using a sphere-selection task and a video frame selection task. Results show that the combination of auditory and visual display outperform both purely auditory and purely visual displays in terms of subjective workload and performance measures.
Chapter 9. Freehand Gesture Interaction

9.1 Freehand Gestures

The increasing number and complexity of novel technologies demands new and innovative interfaces. Free-hand gestures allow natural and intuitive interaction that, in contrast to common devices such as mice or keyboards, permit additional degrees of freedom using the human hand, including wrist rotation and orientation and position and orientation of the fingers to each other. Free-hand gestures are by no means a new approach – however, they have only recently become practical for end users through new developments such as the Leap Motion controller. Stern et al. [139] note that interest and research in free-hand gestures is especially prominent in the fields of medical systems and assistance, entertainment, crisis management, and human-robot interaction.

Even though free-hand gestures are promising, there are challenges facing its application as a standard tool for human-computer interaction. In addition to the fact that gesture recognition is a challenge, even with modern systems, free-hand interaction is often not preferred by users [29]. Free-hand gestures are often unfamiliar; due to a lack of standards, users must learn new gestures [139]. Users are frustrated by the lack of tactile feedback, which is found in almost all conventional forms of input, including mice, keyboards, joysticks, and controllers [113]. Tactile interaction with the aforementioned devices provides instantaneous feedback during operation. In addition, most tactile feedback methods also provide audible sounds when moved or pressed. This problem has been noted in the use of touchscreens; attempts have been made to supplement such screen with tactile [74] and auditory [4] feedback to reduce error rates and improve satisfaction.

In this work, the problem of absent secondary feedback for free-hand gestures is tackled by using auditory display. Three concepts were developed and compared: 1) auditory display with bare-bones visual feedback, 2) solely visual feedback, and 3) combined audiovisual feedback. The evaluation should help determine whether auditory feedback leads to greater speed and accuracy during free-hand gesture interaction, but also whether this leads to improved user experience and acceptance.

Previous investigations into the augmentation of free-hand gestures with feedback primarily focus on visual feedback, tactile feedback, and auditory feedback. Studies that deal with the design of free-hand gestures usually show the user the current gesture mode [73], a 3D model [119] or a live video of the hand [151] so that the user can see the hands and assess whether the gesture was correctly given or recognized by the system. Approaches that
9.2. Method

To determine whether interaction using free-hand gestures can be improved using auditory feedback, three interaction concepts were developed for evaluation using two different evaluation tasks. For gesture recognition, the Leap Motion controller was used in combination with the Unity game development platform to produce a flexible framework for gesture development. To generate real-time sound synthesis, the PureData software environment and a pair of Logitech Z120 multimedia stereo loudspeakers were employed.
9.2.1 Hand Gesture Recognition Framework

Currently, two systems dominate the market for providing relatively simple free-hand gesture recognition: the Leap Motion controller and the Microsoft Kinect. The Leap Motion controller was chosen thanks to its ability to record up to 300 images per second and up to 0.2 mm accuracy in detecting hand and fingers. The Kinect, in contrast, was found to be unsuitable due to lower resolution and accuracy [87]. In addition to the accuracy, the Leap Motion controller offers a mature and robust high-level API for the Unity development platform. The Leap Motion controller uses infrared and can only detect the hand from one side and is therefore prone to occlusion. In addition, the tracking area of the Leap is limited to 25 to 600 mm from the sensor. In pilot studies performed in preparation for this evaluation, a correlation between hand size and accuracy was found. The controller exhibited noticeably lower accuracy when used by participants with smaller hands. The controller is also susceptible to interference from infrared radiation, which limits its tracking quality, for example, under sunlight or halogen lamps, necessitating a testing environment free from direct sunlight or such lamps. The ‘capsule hand’ included in the Leap Motion framework was used to represent the hand in Unity. This model consists of white cylinders and colored spheres at joints.

9.2.2 Auditory Display Framework

PureData [120] is a visual programming environment that allows real-time sound to be produced. To produce auditory feedback, so-called ‘patches’ are fed with signals or data to control synthesizer elements such as oscillators, filters, or noise sources so that the changes in underlying sent data can be transformed into changes in sound and thus be heard by the user. The Open Sound Control protocol [160] was used in a client-server scenario to send data from the hand gesture recognition environment to the PureData synthesizer in real-time.

9.2.3 Interaction Concepts

To investigate the effect of auditory feedback on free-hand gesture interaction, three concepts were developed for two tasks. In Concept A, visual feedback was kept to a minimum: all gestures were supported primarily by auditory feedback. The participant sees only the position of the hand as a sphere. In Concept B, the participant receives enriched visual feedback concerning the completion of gestures and of the task. The hand is represented
as a complete cylinder-and-sphere model. In Concept C, visual and auditory feedback methods are combined. The following hypotheses were derived for these concepts. Concept A (auditory feedback) should perform the weakest of the three in both subjective as well as performance measures, because no complete hand is visible during task completion. Concept B (visual feedback) should perform better than Concept A in both subjective and performance measures, because the hand is visible and errors can be quickly corrected. Concept C should outperform both other concepts, because then hand is visible, and auditory feedback compensates for the loss of tactile feedback.

### 9.2.4 Task 1: Sphere Selection

For Task 1, participants completed an experiment based on ISO 9241-420 [77], for which 8 spheres embedded in a 3D space are oriented in a circle and the participant must select a given sphere. The sphere to be selected was colored orange, and the previously selected spheres were colored green, see Figure 9.1.

![Experimental task 1: The participant has just selected a sphere (green). The next sphere to be selected appears in orange.](image)
To select a sphere, participants were asked to reach into the 3D space with an outstretched hand and close their hand. If so, the sphere was colored blue. If the participant entered with the hand not outstretched, this was considered an error, in which case sphere was colored red and the participant was required to repeat the selection. Participants were asked to select as many spheres as possible with both hands within 30 seconds. The remaining amount of time was represented as a green bar on the left side of the screen. Three feedback concepts were developed for Task 1:

**Concept A**  The participants see no representation of their hands as a 3D model, but rather as a sphere which represents their position. For audio feedback, three auditory icons were produced: a click sound when entering a sphere, an inharmonic error earcon when entering an incorrect sphere, and a harmonic earcon when selecting the correct sphere.

**Concept B**  The participants see a representation of the hands as 3D models. Spheres are colored blue when correctly selected and red when incorrectly selected.

**Concept C**  The participants receive both auditory and visual feedback from Concepts A and B, including 3D hand representation and all 3 auditory icons.
9.2.5 Task 2: Video Navigation

For Task 2, participants completed a more complicated assignment in which a defined image must be selected from a video. This image was shown with a red border, which should be enlarged and shifted until it fit a green border also superimposed on the image. In addition, a horizontal, transparent bar was displayed as a timeline to show current and desired image positions. Three gestures were developed:

**Scroll Gesture** To navigate through the video, the hand should be rotated 90° to the screen directly over the Leap controller. Swiping to the right advances to the next frame in the video, while swiping left advances to the previous frame. The speed of frame change depends on the angle of the hand towards the left or right.

**Shift Gesture** To shift the image, the palm of the hand should be parallel to the monitor, help open. As soon as this gesture is recognized, the frame begins to move relative to the position of the hand.

**Zoom Gesture** To enlarge or reduce the image, the same hand position is used as in the shift gesture, although the hand is to be in a closed fist position. Thus, the participant can ‘grab’ the image and pull out from or push into the monitor.

For each feedback concept, two test runs were completed, each with a different video. As soon as the software recognized the participants’ hands, a timer counted down from 5 to 0. After 0, the task was to be completed as quickly as possibly. For Task 2, three feedback concepts were developed:

**Concept A** The hand is, again, visualized as a sphere. The complexity of the task demanded a more complex auditory display. To acknowledge gesture recognition, a harmonic, 5-note ascending earcon is played back in PureData. To represent the end of a recognized gesture, the same earcon is played in descending order, from the highest to the lowest note. The frame scrolling gesture is implemented as a ‘pulse train’ for which a series of clicks are played back corresponding to each frame change. These are played back in reverse when scrolling backwards through the video. The auditory display for the shift gesture uses model-based sonification to support the participant in recognizing when images are shifted. A series of filtered noise generators gives the auditory effect of rubbing or moving a hard object over a sandy surface. The speed of gesture motion is mapped both to the variation in the speed of
noise generator amplitude modulation as well as the cut-off frequency of a series of low-pass filters, so that faster gesture motions correspond to hearing faster rubbing or motion in the model-based sonification. For the zoom gesture, model-based sonification is used again to support the participant in hearing whether the image is being enlarged or reduced. Here, an approximation of a Shepard [134] tone is used, which gives the illusion of a never-ending increase or decrease in tone pitch. Enlargement is supported by increasing pitch, and reduction by decreasing pitch.

**Concept B** The hand is shown as a 3D representation. The hand is colored blue for scrolling; red for shifting, and yellow for zooming (see Figure 9.3).

**Concept C** In this feedback concept, the preceding two feedback concepts are combined. Thus the participant receives auditory feedback as well as a 3D model of the hands which are colored according to the current gesture.

### 9.2.6 Evaluation

The participants were asked whether they were colorblind or had a hearing impairment. Afterwards, their left hand was photographed for measuring its size. Participants were seated on an adjustable-height chair. For training, a test scene was produced with which participants could see their hand visualized as a 3D model. They received information concerning the operation of the Leap controller, how to avoid hand occlusion and how to stay within the tracking area of the controller. The participants were instructed using a series of slides on how to complete the two tasks and given time (ca. 10 minutes) to familiarize themselves with each feedback concept. Participants were given the opportunity to adjust volume levels to their satisfaction. For
each task, the sequence of feedback concepts A, B, and C were permuted for
each participant to reduce training effects. After one feedback concept was
completed, lasting 2 to 3 minutes, for each task, the participants were asked
to complete both the NASA-TLX [69] and QUESI [154] questionnaires. Accu-
tracy and performance data were gathered using the developed program. As
soon as the sixth and last task was completed and the questionnaires com-
pleted, each participant was asked three questions regarding the auditory
displays and acceptance.

9.3 Results

In total, 14 participants (10 males, 4 females, aged 20-29 yr.) completed the
evaluation, none colorblind or with hearing impairment. User experience
was evaluated with two questionnaires, NASA Raw TLX for subjective work-
load [69] and QUESI for participant satisfaction [154]. These were completed
after each of the 6 tests.

For Task 1, average TLX workload scores were 34.18 for Concept A (au-
ditory display), 35.85 for Concept B (visual display), and 28.33 for Concept
C (combined feedback) on a scale of 0 to 100 where 0 is low workload and
100 is high workload. For Task 2, scores were 39.60 (A), 45.42 (B), and 39.58
(C). For the QUESI questionnaire, average Task 1 scores were 3.96 (A), 3.50
(B), and 3.75 (C), and for Task 2, 3.31 (A), 3.44 (B), and 3.75 (C), where higher
scores indicate increased satisfaction on a scale of 1 to 5. Agreement to the
statement, “The sound bothered me” was 1.64, on a scale of 1 to 5 where 1
indicated “completely disagree” and 5 “completely agree.” For the statement
“The sound helped me” agreement was 4.36. Performance measures for Task
1 included number of spheres correctly selected, incorrectly selected spheres,
and average time between selection. Of interest, the number of incorrectly
selected spheres (error) averaged 1.00 (A), 1.34 (B), and 1.00 (C), and average
times between selection were 1.59 s (A), 1.51 s (B), and 1.26 s (C).

For Task 2, completion time was measured, resulting in average total
task completion times of 48.4 s (A), 49.6 s (B), and 31.6 s (C). Unfortunately,
due to the small number of participants, typical levels of significance could
not be reached, except for the comparison of Task 2 completion time of Con-
cept C with those of Concepts A and B (p = 0.042) and the Task 2 error rate
comparison of Concept C with Concepts A and B (p = 0.034).
9.4 Discussion

This work presents an investigation into the combination of auditory and visual feedback methods for two tasks: a standardized sphere-selection task and a video frame selection, enlargement, and shifting task. A user study was completed with 14 participants. Results suggest that the combination of auditory and visual feedback provide lower subjective workload, lower task completion time, and fewer errors than either solely auditory or solely visual feedback. Free-hand gestures are not yet a standard input method for human-computer interaction. The number of degrees of freedom of the human hand and the lack of standards for gesture-based interaction methods provide substantial hindrances to its use. Designing auditory display is also not trivial; both psychoacoustic properties of human hearing as well as aesthetic concerns are burdens that prohibit a general implementation of complex auditory display as a standard means of feedback. Thanks to technological breakthroughs such as the Leap Motion controller as well as real-time sound synthesis software, the combination of both fields of research has become more approachable. However, investigations into the combination of novel gesture recognition accompanied by auditory display is sparse, but could increase in coming years through intensified focus in the field of augmented and virtual reality. Especially in medical contexts, where sterile interaction is paramount, research into gesture interaction is especially prudent.

Even though comparisons between concepts for many of the measures could not reach a typical level significance, free-hand gestures supported by auditory display appears to be a promising combination. Because of the lack of secondary tactile feedback in free-hand interaction, auditory display is essential in knowing when gestures are recognized and whether the recognized gesture is correct. Future work should especially focus on model-based auditory display in which the user receives continuous feedback for free-hand gestures. In this way, a user could ‘play’ the gestures similar to playing an instrument, thereby providing a finer degree of control compared to simple warning and alarm sounds that trigger when a certain action has been executed. This type of feedback was shown in Park et al. [113] to have the highest performance, and in this work showed significant improvements in task performance time for Task 2.
9.5 Conclusion

This work describes three concept for feedback for free-hand gesture interaction with a Leap Motion controller. Auditory, visual, and combined audiovisual feedback methods were developed to support the user in evaluating two separate screen-based tasks. Although in some cases a typical level of significance could not be reached, results show that combined audiovisual display outperforms and is preferable to purely auditory or purely visual displays. The results suggest that increased focus into the multimodal effects of auditory and visual feedback for hand gestures is warranted, especially due to the lack of secondary tactile feedback inherent in typical free-hand interaction.
Chapter 10

Conclusion and Outlook

The increasing presence of both image-guidance systems in the operating room and the evident potential of touchless interaction systems have led to the development of clinical systems for medical personnel that provide great benefit in terms of information access. This thesis explores the use of auditory display to improve interaction with these systems. Results from a range of evaluations based on clinical use cases show that the use of auditory display in the operating room for both guidance as well as touchless interaction show great promise and warrants increased attention in this exciting and growing field.
10.1 Challenges

During image-guided interventions, clinicians can access the preoperatively or intraoperatively acquired imaging to help increase orientation and successfully follow an interventional strategy. However, existing visual methods are often overloaded and complex, and require clinicians to often switch views between patient and screen, thereby increasing workload and reducing concentration on the patient. For novel sterile touchless interaction approaches for the operating room, difficulties lie primarily in the loss of secondary tactile feedback that is present when interacting with conventional input devices such as keyboards and mice.

A great difficulty encountered when exploring the use of auditory display for the operating room was receiving the essential access to face-to-face encounters with medical professionals and clinics, which is essential for the contextual inquiry that helps determine current difficulties with image guidance and other operating room systems that are faced by practicing clinicians. Clinicians are continually pressed for time, and interviews or clinical observation visits were thus limited. Therefore, in addition to a small but crucial number of personal consultation with medical professionals, much of the contextual inquiry for each project was limited to reviews of literature as well as discussions with partner medical technology partners. This, despite a degree of access that allowed very fruitful collaboration and research to take place and generate significant results, it was insufficient to answer overarching questions of applicability of auditory display on a greater scale.

Each of the projects clearly demonstrates that auditory display can mitigate certain deficiencies in current image-guidance and operating room systems, but due to the restrictions of the scale of the partnerships with other institutions, only singular, individual auditory displays could be evaluated in a laboratory setting compared to the control of existing visual information systems. The logistics of the partnerships meant that comparing, for instance, multiple auditory displays that explored the benefits of alternative mappings for each scenario was not possible.

10.2 Contributions of This thesis

This thesis concentrates on researching methods that aim to improve access to preoperative and intraoperative information by using novel auditory display. Chapter 2 provides a thorough review of the state of the art of the use of auditory display in image-guided interventions. Chapters 3 through 7 investigate the use of novel auditory display to aid clinicians when performing
image-guided interventions. These primarily concentrate on guided, tracked instrument placement ranging across a variety of clinical contexts. Chapters 8 and 9 explore the use auditory display to augment touchless interaction systems for the operating room.

**Chapter 3** focuses on needle placement for radiological ablation needle guidance. An auditory display using alternating tones was developed to support insertion of a needle, helping guide the user to place tip and handle correctly. This study describes the first attempt in the literature to evaluate ‘blind’ placement of a needle, and results showed that the auditory display was indeed capable of screen-free guidance. Accuracy of placement using solely auditory display approached that placement using a screen, although task time was significantly higher. Using combined audiovisual display increased accuracy compared to visual display and reduced time viewing the screen.

**Chapter 4** describes auditory display to support navigation of a teleoperated continuum robot in the tunnel-like scenario of the transnasal passage to the sphenoidal sinus. An auditory display that features sung syllables was investigated for three dimensions of guidance. Results showed that when using auditory display, users performed significantly more efficiently, and using the auditory display was rated with significantly lower overall subjective workload and significantly higher usability and satisfaction than using only the visual display.

**Chapter 5** investigates the use of mixed-reality navigation for laparoscopy tasks. As part of the larger novel integrated system, an immersive auditory display implementing tone comparison and stereo positioning was designed to provide the user with three-dimensional position cues to aid guidance during intraoperative laparoscope navigation. Results of an evaluation on peg-transfer tasks using the mixed reality system revealed longer task completion times compared to a standard navigation system for simple tasks, but faster times, increased accuracy, and decreased workload for more complex tasks.

**Chapter 6** describes an auditory display for open liver surgery to guide a tracked dissector towards and remain on a preoperatively defined resection line. The auditory display maps the distance of the tip of the instrument to the resection line divided into 3 margins, informing when either inside a safe zone, a warning zone, or outside the warning zone. An evaluation with clinicians showed that when using a combined audiovisual display, participants exhibited increased accuracy and time spent looking at the patient, and task completion time.
Chapter 7 presents an auditory display for communicating measured fluorescence intensity values of a handheld probe directly to the surgeon during open brain surgery. The auditory display that implements earcons to represent three intensity levels and an error state was evaluated in terms of response accuracy and response latency. After training, almost all responses by participants were correct, and response latency was low enough to warrant further investigation in the operating room to eliminate the need for an assistant to verbally relay intensity levels shown on a screen and play them directly to a clinician.

Chapter 8 explores the use of auditory display to augment interaction with an eye-tracking input method for sterile operating rooms. The auditory display, which consists of earcons and a parameter mapping model, was implemented to replace the lost sense of touch that is removed when interacting with touchless systems such as eye tracking. In an evaluation based on typical scrub nurse tasks, using auditory display reduced reaction times while improving subjective measures, including usefulness, user satisfaction, and cognitive workload compared to interaction without auditory display.

Chapter 9 similarly presents an auditory display method to augment sterile interaction in the operating room. A system for interacting with images using freehand gestures captured by a Leap motion controller was augmented with an auditory display implementing parameter mapping models, including an approximation of a Shepard tone as well as auditory icons simulating rubbing or moving objects on a surface. An evaluation showed that the combination of auditory and visual feedback provide lower subjective workload, lower task completion time, and fewer errors than either solely auditory or solely visual feedback.

In all, the thesis addresses a substantial range of use challenges drawn from use cases gathered from the clinical routine, and explores auditory display spanning a multitude of techniques for 1D, 2D, and 3D navigation, as well as fluorescence intensity values and augmentation for touchless interaction. Across all chapters, the primary benefits for auditory or combined audiovisual display compared to visual-only display include increases in accuracy, decreases in cognitive workload, and increases in usability, whereas drawbacks included increases in task completion time and in some cases, increase in cognitive workload.
10.3 Proposals for Future Research

This thesis poses many more questions than it answers. Although it shows substantial progress in the establishing the importance and promise of auditory display in the operating room, it is only a first step in understanding its role and potential. The overarching goal of future research should be the integration of the body of knowledge gained from the promising results of individual projects described in each chapter of this thesis.

Future research should expand the current state of the art in both breadth, across additional clinical use cases, as well as depth, to further develop and validate promising method of auditory display. Although this thesis has shown that auditory display has provided significant benefit with respect to both qualitative and quantitative measures, an extension of the research should provide increased opportunities to also compare many different parameter mappings with each other and against control groups using existing visually based interventional systems. Based upon knowledge gathered from individual evaluations, a generally applicable set of context-sensitive audiovisual display methods could form the basis for the compilation of a catalog of auditory display for optimal delivery pre- and intraoperative data to the clinician during a much wider variety of tasks. Creating a generally applicable catalog of audiovisual display methods would accomplish two goals. First, similar methods would be available to surgeons across multiple tasks, reducing confusion and accelerating training. Second, this development could help answer fundamental questions raised in the future contextual inquiries concerning the best way of augmenting clinicians’ interaction with operating room systems.

In addition to expanding into additional clinical applications and the generation of a generally applicable catalog of auditory display methods, future research should further explore questions of multimodality in new methods. There is a paucity of knowledge concerning the intricacies of multimodal operation between auditory and visual displays, not to mention in clinical settings. This thesis provides some insight into the optimal modalities needed by the clinician for each stage of the described applications, but does not explore this factor in sufficient depth. Future research should not only determine whether auditory display is preferable compared to visual display for a certain clinical use case, but also determine the exact moments in which each modality is suitable, and in which cases a combined display is desired. Context sensitivity is of great importance to this goal, and future hybrid audiovisual display systems could rely more heavily on surgical process modeling and intelligent operating room technology [50] for tighter integration with ongoing clinical tasks. In addition to a higher degree of harmony
between auditory and visual displays, future research must bring auditory display into a greater degree of contact with novel interaction possibilities such as foot control [49, 70], myoelectric (muscle-based) gesture control [73], and other touchless input methods [106].

After validating additional methods based on the applications discussed, future research should intensify cooperation with clinicians and clinics. Although exchange with clinicians was a significant part of this thesis, it was insufficient to validate the use of the explored auditory display methods for clinical use. This can only be accomplished with a greater integration of efforts at all stages of research, from preliminary contextual inquiry to prototype development to evaluation. The latter is of utmost importance: a clinical evaluation is the ultimate factor in deciding the future of auditory display for operating room systems.
Author’s Publications

Journal Publications


Conference Proceedings


David Black, Sebastian Heise, Jörn Loviscach. “Generic Sound Effects to Aid in Audio Retrieval”. In: *Proceedings of Audio Engineering Society*, Munich, Germany, 2009.

### List of Figures

2.1 Review of approaches to auditory display mapping .......................... 15

3.1 Needle placement visualization ................................................. 31
3.2 Test setup for needle placement task ........................................ 32
3.3 Auditory display method for needle placement .............................. 33
3.4 Phantom and tracked needle .................................................... 34
3.5 Placement accuracy ............................................................. 38
3.6 Task completion time .......................................................... 38
3.7 Subjective workload ............................................................ 39

4.1 Teleoperation of tubular continuum robots ................................. 47
4.2 Continuum robot simulation environment .................................... 49
4.3 Distance calculation for robot end effector ................................. 50
4.4 Auditory display method for continuum robot navigation .............. 51
4.5 Evaluation setup .................................................................. 54
4.6 Performance results for error, task time, and efficiency ............... 57
4.7 Questionnaire results: NASA TLX and van der Laan ................... 59
4.8 Overall ranking results ......................................................... 61

5.1 CAD model and 3D printed attachment on Oculus Rift .................. 67
5.2 Mixed reality navigation software layout .................................... 68
5.3 User performing task with mixed reality setup ............................... 72
5.4 Example trajectory for grasper ................................................ 73
5.5 Example trajectory for mixed reality trial .................................... 74

6.1 Liver phantom with floral foam (green) inside .............................. 85
6.2 Liver resection evaluation setup .............................................. 86
6.3 Surgeon glancing at the navigation system screen ....................... 86
6.4 Proportion of time spent looking at navigation screen ................. 88
6.5 Distance between planned and subject-drawn resection lines ......... 89
6.6 Mean time per task for combined and visual condition ............... 90

7.1 Fluorescence guided surgery scenario ....................................... 94
7.2 Fluorescence spectra measured during surgery ............................ 95
7.3 Phantom setup for fluorescence auditory display ........................ 97
7.4 Phantom well arrangement ..................................................... 98
7.5 Flowchart of signal generation in LabVIEW ............................... 99
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.6</td>
<td>Profile of the auditory display</td>
<td>101</td>
</tr>
<tr>
<td>7.7</td>
<td>Median of latency versus accuracy for each participant</td>
<td>107</td>
</tr>
<tr>
<td>7.8</td>
<td>Latency variations for each participant</td>
<td>108</td>
</tr>
<tr>
<td>8.1</td>
<td>Eye-tracking setup in laboratory operating room</td>
<td>115</td>
</tr>
<tr>
<td>8.2</td>
<td>Eye tracking QR code example</td>
<td>117</td>
</tr>
<tr>
<td>8.3</td>
<td>Auditory display for gaze gestures</td>
<td>119</td>
</tr>
<tr>
<td>8.4</td>
<td>Laboratory setup for eye-tracking evaluation</td>
<td>120</td>
</tr>
<tr>
<td>8.5</td>
<td>Reaction time results</td>
<td>125</td>
</tr>
<tr>
<td>9.1</td>
<td>Freehand task 1: sphere selected</td>
<td>135</td>
</tr>
<tr>
<td>9.2</td>
<td>Freehand task 1: entering sphere, incorrect sphere</td>
<td>136</td>
</tr>
<tr>
<td>9.3</td>
<td>Freehand task 2: scroll, shift, and zoom gestures</td>
<td>138</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Clinical applications and aud. disp. using proximity alerts . . . . 11
2.2 Clinical applications and continuous auditory display . . . . 11
2.3 Informal and laboratory evaluations in literature . . . . . . . 19
2.4 Clinical evaluations in literature . . . . . . . . . . . . . . . . . . 20
3.1 Summary of statistical analysis . . . . . . . . . . . . . . . . . 36
4.1 Agreement question results . . . . . . . . . . . . . . . . . . . 60
6.1 Resection guidance questionnaire results . . . . . . . . . . . . . 91
7.1 Fluorescence phantom composition . . . . . . . . . . . . . . . . 100
7.2 Confusion matrix for intuition test results . . . . . . . . . . . . 106
7.3 Confusion matrix for function response accuracy . . . . . . . . 106
8.1 Summary of statistical results for eye tracking . . . . . . . . . 124
8.2 Eyetracking subjective questionnaire results . . . . . . . . . . . 126
Bibliography


