

POLITECNICO
MILANO 1863

Department of Electronics, Information and Bioengineering
Doctoral Programme In Information Technology

Physics-based Extended Reality for Real-time Guidance in Minimally Invasive Cardiovascular Interventions

Doctoral Dissertation of:
Veronica Ruozzi

Advisor: Prof. Emiliano Votta
Co-advisors: Riccardo Vismara
Tutor: Prof. Luigi La Barbera
Year 2025 - XXXVIII Cycle

Abstract

Percutaneous interventions have transformed surgical practice by enabling minimally invasive access to anatomical structures. While surgical planning based on preoperative imaging and 3D anatomical reconstruction supports procedure design, intraoperative guidance remains intrinsically challenging. Surgeons must infer tool position without direct visualization, relying on intraoperative imaging that provides only partial, modality-specific information, while haptic feedback through long, flexible instruments is highly attenuated. In cardiovascular percutaneous procedures, these challenges are amplified by the fast and continuous motion of cardiac targets and the deformability of cardiovascular and surrounding soft tissues. These factors demand highly time-resolved feedback on tool-target relative pose and updated 3D reconstructions accounting for intraoperative tissue deformations.

The present thesis introduces physics-based computational tools to address these needs in cardiovascular percutaneous procedures. It proposes finite element modeling to compute tissue deformations invisible to X-ray fluoroscopy, and sonification to provide feedback on tool and cardiac target pose, both integrated within an Extended Reality framework. The developed solutions are showcased through three use cases: (i) transcatheter navigation where catheter sensing and biomechanical simulation predict vessel deformations; (ii) percutaneous cardiac access where physics-driven sonification encodes needle proximity to moving structures; and (iii) a generalizable framework formalizing biomechanical-acoustic coupling for tool-tissue interaction representation across surgical domains.

This work advances computer-assisted surgery by establishing validated methodologies for physics-based surgical interaction modeling, demonstrating that multisensory Extended Reality interfaces effectively address limitations of purely visual guidance. These contributions establish a foundation for surgical navigation systems where imaging, biomechanics, and sensory feedback function as coupled elements of a unified computational framework for real-time intervention support.

Keywords: Surgical Interaction Modeling, Medical Extended Reality, Physics-driven Sonification, Percutaneous Cardiovascular Interventions

Abstract in lingua italiana

Gli interventi percutanei hanno rivoluzionato la pratica chirurgica, consentendo l'accesso mini-invasivo alle strutture anatomiche. Sebbene la pianificazione basata su imaging pre-operatorio e ricostruzioni anatomiche 3D supporti la pianificazione della procedura, la guida intraoperatoria rimane intrinsecamente complessa. I chirurghi devono dedurre la posizione dello strumento in assenza di visualizzazione diretta, affidandosi a immagini intraoperatorie che forniscono informazioni parziali, mentre il feedback tattile attraverso strumenti lunghi e flessibili è fortemente attenuato. Nelle procedure cardiovascolari percutanee, queste difficoltà sono amplificate dal movimento rapido e continuo dei target cardiaci e dalla deformabilità dei tessuti circostanti e cardiovascolari stessi. Questi fattori richiedono un feedback temporale altamente accurato sulla posizione relativa tra strumento e bersaglio, nonché aggiornamenti delle ricostruzioni 3D che tengano conto delle deformazioni intraoperatorie.

La presente tesi introduce strumenti computazionali basati sulla fisica per rispondere a queste esigenze. In particolare, propone l'uso della modellazione a elementi finiti per stimare le deformazioni dei tessuti non visibili alla fluoroscopia e della sonificazione per fornire un feedback sulla posizione relativa tra strumento e bersaglio cardiaco dinamico, entrambi integrati in un framework di realtà estesa. Le soluzioni sviluppate sono illustrate attraverso tre casi d'uso: (i) la navigazione transcatetere, in cui i dati di un catetere sensorizzato, integrati con simulazione biomeccanica, consentono la predizione delle deformazioni vascolari; (ii) l'accesso cardiaco percutaneo, in cui la sonificazione basata sulla fisica codifica la vicinanza dell'ago a strutture in movimento; e (iii) un paradigma generale che integra biomeccanica e sintesi acustica per rappresentare l'interazione strumento–tessuto in diversi domini chirurgici.

Questa tesi rappresenta un progresso significativo nell'ambito della computer-assisted surgery, proponendo metodologie validate per la modellizzazione dell'interazione chirurgica. I risultati dimostrano che le interfacce multisensoriali di realtà estesa possono superare i limiti dei modelli di guida puramente visiva, aprendo nuove prospettive per la navigazione negli interventi percutanei attraverso l'integrazione di imaging, biomeccanica e feedback multi-sensoriale.

Parole chiave: Modellazione dell'Interazione Chirurgica, Realtà Estesa, Sonificazione, Interventi Cardiovascolari Percutanei

Contents

Abstract **i**

Abstract in lingua italiana **iii**

Contents **v**

1 Introduction **1**
1.1 Aim and Outline of the Thesis 5

2 Theoretical Background and State of the Art **7**
2.1 Computer-Assisted Surgery Foundations 7
2.1.1 Medical Image Processing 8
2.1.2 Intraoperative Tool Tracking 13
2.1.3 Registration Approaches 16
2.2 Medical Extended Reality (XR) 19
2.2.1 Definitions and Technology 19
2.2.2 Medical Extended Reality Systems 21
2.2.3 Registration for XR-Guided Surgery 23
2.2.4 Promising Clinical Applications 24
2.2.5 Intraoperative Guidance in Cardiovascular Procedures 25
2.3 Anatomical Updating Strategies for Surgical Navigation 26
2.3.1 Image-Based Approaches 28
2.3.2 Interactive Physics-based Modeling 28
2.4 Medical Sonification 32
2.4.1 Visual Channel Limitations and Multisensory Perception 32
2.4.2 Parameter-Mapping Approaches 33
2.4.3 Physics-Driven Model-Based Approaches 34

3 Motivation and Fundamental Impact **37**

3.1 Main Motivation . . . 37
3.2 Surgical Interaction Modeling . . . 38

4 Enhanced Catheter Monitoring in Transcatheter Procedures 41
4.1 Biomechanical Modeling of Catheter-Vessel Interaction . . . 42
4.2 Contact Modeling Approaches . . . 44
4.2.1 Lagrange Multiplier Formulation . . . 45
4.2.2 Computational Cost Analysis . . . 47
4.3 Methods . . . 50
4.3.1 In-Vitro Setup . . . 53
4.3.2 Initialization . . . 53
4.3.3 Registration . . . 55
4.3.4 Runtime . . . 55
4.3.5 Computational Benchmarking . . . 56
4.3.6 In-vitro Experiments . . . 60
4.4 Results . . . 61
4.4.1 Interactive Performance . . . 61
4.4.2 Accuracy of Simulated Vessel Deformation . . . 62
4.5 Discussion . . . 65
4.5.1 Novelty and Technical Contributions . . . 65
4.5.2 Reliability and Performance . . . 67
4.6 Limitations and Future Directions . . . 69
4.7 Final Remarks . . . 70

5 Multisensory Navigation for Percutaneous Cardiac Access 71
5.1 Subxiphoid Percutaneous Epicardial Access . . . 72
5.1.1 Proposed Framework . . . 75
5.2 Methods . . . 75
5.2.1 Dynamic Anatomy Reconstruction and Trajectory Planning for PEA 76
5.2.2 Real-Time Tool Navigation . . . 78
5.2.3 Multisensory Extended Reality Surgical Navigation Display . . . 78
5.2.4 System Evaluation: The Phantom Study . . . 85
5.3 Results . . . 89
5.3.1 Navigation Task Outcome Analysis . . . 90
5.3.2 Needle Tip Placement Analysis . . . 91
5.3.3 Execution Time . . . 96
5.3.4 Workload Assessment . . . 97
5.3.5 Experts' Subjective Feedback . . . 97

5.4 Discussion . . . 97
5.5 Limitations and Future Directions . . . 99
5.6 Final Remarks . . . 102

6 BioSonix: A New Paradigm for Perceptually Rendering Tissue Deformations from Tool Interactions 103
6.1 Perceptual Challenge of Tool-Tissue Interaction . . . 104
6.1.1 Proposed Solution . . . 105
6.2 BioSonix Design Framework . . . 106
6.2.1 Anatomical Domain Module . . . 107
6.2.2 Mapping Module . . . 108
6.2.3 Sonification Domain Module . . . 109
6.3 Methods . . . 109
6.3.1 Abstract Anatomical Model . . . 110
6.3.2 Sound Model . . . 111
6.3.3 Model Optimization . . . 112
6.3.4 Fine Tuninig . . . 113
6.4 Experiments and Results . . . 113
6.4.1 Toward Realistic Models . . . 114
6.4.2 Real Anatomical Models . . . 115
6.5 User Experiments and Expert Evaluation . . . 115
6.5.1 General Perceptibility Study . . . 115
6.5.2 Expert Evaluation of the Needle Biopsy Task . . . 116
6.6 Discussion . . . 117
6.7 Limitations and Future Directions . . . 118
6.8 Final Remarks . . . 119

7 Conclusions and Future Directions 121
7.1 Thesis Contribution . . . 121
7.2 General Conclusions . . . 123
7.2.1 Translating Research into Clinical Practice . . . 124
7.3 Future Directions . . . 125

Bibliography 129

A Appendix A 153

B Cardiac Dynamic Reconstruction: Technical Background **155**
B.1 4-Dimensional Computed Tomography Imaging 155
B.1.1 ECG Gating 156
B.2 Deformable Image Registration 156
B.2.1 Validation of Deformable Image Registration 157
B.3 Patient-Specific 4DCT Dataset Acquisition 159
B.4 Implementation of Deformable Image Registration for Pericardial Segmentation 159
B.4.1 Validation of the Registration Framework 160

List of Figures **161**

List of Tables **167**

Acknowledgements **169**

List of Acronyms **171**

List of Acronyms **171**

1 | Introduction

The Operating Room (OR) is a complex, variable and demanding environment: surgeons, anesthesiologists, nurses, and specialists from multiple disciplines coordinate complex tasks while managing advanced technologies, from intraoperative imaging systems to robotic platforms, all under significant time pressure and stress [111]. This complexity is further amplified in Percutaneous Interventions (PIs), which have progressively replaced traditional open-sky surgery [17, 21, 57]. These techniques represent a paradigm shift from direct visual access and fully exposed operative fields to indirect access to anatomical targets achieved through the navigation of guidewires, catheters, and needles through the patient's body.

The cardiovascular field exemplifies this transformation: cardiovascular PIs now treat intracardiac or intravascular conditions that previously required open-heart surgery or thoracotomy. Over the past two decades, catheter-based interventions for Structural Heart Disease (SHD) have experienced exponential growth, becoming established alternatives to surgical approaches [109, 227], particularly for high-risk patients. In these procedures, catheters are introduced through peripheral vessels and navigated along the vascular bed to reach targeted cardiac structures where therapeutic devices are deployed. Examples include Transcatheter Aortic Valve Replacement (TAVR), mitral valve repair using MitraClip or Transcatheter Edge-to-edge Repair (TEER), and Percutaneous Left Atrial Appendage Closure (PLAAC). In parallel, needle-based procedures enable direct access to cardiac structures. For instance, Percutaneous Epicardial Access (PEA) for ventricular tachycardia ablation allows electrophysiological mapping of the myocardial substrate through direct needle insertion below the sternum [11].

In PIs, interventional tools such as catheters, guidewires, and needles are manipulated from a proximal handle or, in more complex delivery systems for transcatheter procedures, through multi-degree-of-freedom knob controls. Because there is no direct line of sight to the operative target, this configuration introduces a fundamental *sensorimotor* decoupling [83]: the operator's actions at the proximal handle are spatially and mechanically separated from their distal consequences at the patient's anatomy. As a result, surgeons

must infer the tool's position and orientation based on prior anatomical knowledge, supplemented by intraoperative imaging and the limited haptic feedback available.

Intraoperative imaging systems used for procedural guidance are constrained by the inherent limitations of their sensing modalities, typically providing only partial and modality-specific information. X-ray fluoroscopy, for example, offers excellent visualization of radiopaque instruments such as guidewires and catheters, but it does not depict surrounding soft tissues or their deformation during manipulation. In contrast, ultrasound provides real-time imaging of soft tissue structures and dynamic anatomical changes, yet tracking metallic instruments remains unreliable due to its narrow field of view and acoustic impedance mismatches that generate strong reflections, shadowing, and other artifacts. These limitations are particularly relevant in cardiovascular PIs, where the anatomy is highly dynamic and tool-tissue interactions are subtle and transient. As a result, multiple complementary imaging modalities, most commonly fluoroscopy and ultrasound, are used concurrently to compensate for the shortcomings of each single imaging technology [198]. Within this multimodal visual environment, the surgeon must continuously perform cognitive integration: mentally aligning prior anatomical knowledge with the live intraoperative scene while synthesizing fragmented image information distributed across several displays, each corresponding to a different modality. This workflow requires frequent shifts of attention between the operative field and multiple external monitors, all while simultaneously manipulating interventional instruments.

Moreover, the mechanical resistance perceived at the proximal handle of an instrument is often only weakly correlated with the events occurring at the distal tool tip [223]. This decoupling is particularly pronounced in long, flexible instruments such as catheters and guidewires. Under these conditions, surgeons have no direct means of perceiving tissue resistance or texture. Instead, they receive only indirect and highly attenuated cues arising from frictional forces, shaft deformation, and tool-tissue interactions transmitted along the device. These cues are heavily damped, producing haptic information that is ambiguous, low-fidelity, and difficult to interpret reliably. This limitation becomes especially critical in procedures where subtle haptic cues inform key clinical decisions, such as detecting tissue penetration during needle advancement in PEA or ensuring adequate and stable contact force in catheter-based interventions, for example during cardiac mapping for arrhythmia ablation or in complex coronary procedures. Unfortunately, despite their clinical importance, the reliable interpretation of these subtle tactile signals remains difficult and strongly dependent on operator expertise [19].

Over the past decades, substantial advances in research and technology have emerged with the aim of addressing these challenges and supporting the growing complexity of emerging

PIs. As a result, modern ORs have evolved into highly data-rich environments, increasingly equipped with advanced sensing technologies, user interfaces, robotic platforms, and integrated image- and data-processing pipelines [102, 189].

At the core of this development, Computer-Assisted Surgery (CAS) has been established as the discipline providing computational tools to support the surgical workflow, from preoperative planning to intraoperative guidance [63].

Its foundations lie in medical image processing: segmentation and three-dimensional (3D) anatomical modeling transform preoperative imaging data, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), into patient-specific representations that enable trajectory planning, risk assessment, and procedural simulation [185]. To ensure continuity between planning and execution, image registration pipelines align preoperative models with intraoperative imaging across different modalities [37]. In cardiovascular PIs, such registration techniques, have demonstrated clinical benefits including improved instrument guidance, reduced radiation exposure, and more reliable target localization [198].

Modern CAS further leverages Artificial Intelligence (AI) and computer vision to automate segmentation, extract anatomical landmarks in real-time, and semantically interpret image sequences, supporting intraoperative decision-making [32, 132, 214]. Complementing these computational foundations, advanced tracking methodologies [172] enable complete surgical navigation systems that localize instruments relative to co-registered imaging and anatomical models.

Despite these advances, guidance information remains predominantly confined to external flat monitors, limiting spatial intuitiveness and requiring surgeons to shift focus away from the operative field. As a result, constraints related to hand-eye coordination, spatial awareness, and cognitive workload inherent to PIs remain largely unresolved, underscoring the need for interfaces capable of delivering real-time, context-aware information directly within the surgeon's natural line of sight. While major progress has been achieved in data processing and algorithmic methodologies, an equally critical challenge lies in understanding how this computational output is conveyed to the surgeon, taking into account user needs, perceptual constraints, and the complexity of the modern OR environment.

Medical Extended Reality (XR) system directly address this limitation by overlaying computer-generated information, like anatomical structures, planned trajectories, and instrument positions, onto the operative scene, improving spatial understanding and ergonomics without obstructing the field [157].

Their application as surgical navigation system has been most thoroughly investigated in domains involving rigid or quasi-rigid anatomical structures, such as neurosurgical, orthopedic, and oral procedures, where stable anatomical landmarks enable preoperative navigational cues to remain spatially accurate throughout the intervention [130]. When deployed as surgical navigation systems in soft-tissue environments, however, a key challenge arises: preoperative anatomical models progressively lose accuracy as the procedure advances, due to tissue deformation, instrument-tissue interactions, and patient motion. Addressing this limitation, several model-updating strategies have been proposed, spanning biomechanical simulation, statistical shape models, and learning-based approaches [72], as well as image-based intraoperative model updating [36]. Nevertheless, these solutions have so far been validated mostly in abdominal settings, where soft-tissue deformation evolves slowly and can be at least partially observed through intraoperative ultrasound, stereo, or depth imaging.

The extension of medical XR system as surgical navigation systems to cardiovascular PIs remains largely unexplored. Unlike the relatively stable anatomies encountered in neurosurgical or orthopedic procedures, cardiovascular structures are comprised of soft tissue that deforms under instrument manipulation, with only partial and fragmented observational data from intraoperative imaging. Moreover, the heart undergoes continuous cyclic displacement and deformation due to cardiac contraction and respiratory motion, which demands guidance information delivered at high temporal resolution.

Taken together, these constraints define surgical navigation in the cardiovascular field as a fundamental *spatiotemporal* problem: providing reliable spatial information requires updating strategies to represent tissue deformation, while the temporal dynamics demand that such updates are conveyed to the surgeon in real-time to preserve guidance validity. This dual requirement raises fundamental questions regarding computational feasibility, appropriate modeling strategies, and effective information delivery.

Regarding the effective information delivery, the inherent dynamics of cardiovascular anatomy raises a question: whether visual augmentation alone, however accurate, is sufficient to convey time-critical navigational information in an environment where the surgeon's visual attention is already heavily committed [46]. This observation motivates the exploration of complementary sensory channels, not as a complete replacement for visual guidance, but as a means of extending its reach into dimensions that vision alone cannot efficiently cover.

Sonification, the encoding of data through structured auditory signals, offers a promising avenue in this regard. Sound can convey multidimensional, time-varying information

through frequency, amplitude, and spectral patterns, exploiting the remarkable temporal sensitivity of the human auditory system. Its omnidirectionality enables continuous perception independent of gaze direction, making it a natural complement to visual guidance in dynamic, high-attention environments. Prior work in the surgical domain has demonstrated its value in improving situational awareness [137] and supporting interventional tasks such as tool navigation [20, 138].

Whether and how these principles can be extended to the cardiovascular domain remains an open question.

Collectively, CAS and medical XR systems form a digital continuum from preoperative modeling to intraoperative execution, and substantial progress has been made along each stage of this pipeline. Sonification has shown promise in surgical navigation tasks, and its properties for conveying time-critical information appear well-suited to the *spatiotemporal* challenges posed by the cardiovascular domain. Yet their integration into a multisensory surgical navigation system surfaces a set of intertwined challenges in modeling and perceptual delivery that remain unexplored.

1.1. Aim and Outline of the Thesis

This thesis investigates (i) the development of **physics-based modeling strategies** and (ii) the design of **multisensory display** that address the spatial and temporal requirements outlined above, enabling the implementation of multisensory surgical navigation systems for cardiovascular PIs.

The approach integrates different technologies: CAS as the computational foundation, medical XR system for sensory augmentation through Optical See-Through Head-mounted Display (OST-HMD) for visual feedback, and sonification for auditory feedback.

This multisensory framework is investigated across two clinically distinct use cases: transcatheter endovascular procedures and needle-based direct cardiac access.

To this end, the thesis is structured as follows. **Chapter 2** establishes the theoretical background and reviews the state of the art, providing the foundation necessary to understand and critically assess the proposed methodologies. **Chapter 3** frames the motivation and the specific impact of this thesis within the reviewed literature.

Chapters 4 and 5 progressively address these methodological challenges in the context of transcatheter endovascular navigation and needle-based intracardiac access, respectively. **Chapter 6** extends these contributions by establishing a framework that directly

couples physics-based modeling with acoustic synthesis, achieving deeper integration between anatomical representation and perceptual delivery and establishing design principles transferable beyond the cardiovascular domain.

Finally, **Chapter 7** synthesizes the findings, discusses their broader implications, and outlines directions for future research.

2 | Theoretical Background and State of the Art

This chapter provides the theoretical background and reviews the state of the art in the technologies and methodologies most relevant to this thesis. This review is also intended to contextualize the clinical scalability and applicability of the contributions presented in subsequent chapters. Special emphasis is placed on the cardiovascular domain, and the specific limitations that characterize this field are analyzed and summarized.

Section 2.1 outlines the core CAS methodologies the developed approaches rely on, covering medical imaging and image processing, intraoperative tool tracking, and advanced registration approaches.

Section 2.2 defines Medical XR in its technological and applicative dimensions, and discusses its integration within CAS pipelines through clinically relevant examples in image-guided surgery and PIs.

Section 2.3 reviews anatomical updating approaches relevant to XR-based surgical navigation, addressing the challenge of accounting for soft tissue deformation. Each method is discussed with respect to the two fundamental requirements of intraoperative use: accuracy and computational interactivity.

Section 2.4 introduces the concept of Medical Sonification as a component for multisensory XR displays, addressing the challenge of effective information delivery. The section illustrates the role of sound as a physical phenomenon and clarifies the importance of physics-based modeling in the design of effective auditory guidance cues.

2.1. Computer-Assisted Surgery Foundations

In this thesis, medical image processing, tracking, and registration are considered enabling technologies for the development of surgical navigation system, rooted in CAS framework.

In 2020, a consensus definition of CAS was produced by experts in the field, describing it

as *"the broad use of information technology frameworks to enhance physicians' skills and augment senses, as in image-guided surgery, cognition, through deep learning and machine learning, and execution, via mechatronic, imaging, and surgical robotics systems, with the aim of providing more precise and safer procedures"* [63].

From a conceptual standpoint, CAS systems can be interpreted as pipelines that integrate patient-specific data, computational models, and intraoperative sensing into a unified framework supporting clinical guidance and decision-making.

Figure 2.1 illustrates the main components of CAS: medical image acquisition and processing yield anatomically meaningful representations; in the intraoperative phase, tracking systems and registration strategies establish a shared spatial reference among heterogeneous data sources, thereby enabling surgical navigation.

Given the breadth and interdisciplinary nature of CAS, this section does not attempt a comprehensive survey of all existing methodologies, but rather focuses on the fundamental components that underpin the approaches developed in this thesis. Their underlying principles are reviewed for contextual completeness, as the studies presented subsequently rely on established methods in these domains rather than introducing novel contributions to them.

2.1.1. Medical Image Processing

Medical image processing is the foundational layer of any CAS pipeline. The ability to process, and interpret patient-specific anatomical information is a prerequisite for planning, guidance, and intraoperative decision-making.

In the clinical setting, imaging modalities can be broadly categorized according to the stage of the workflow in which they are primarily employed: *preoperative* modalities, aimed at diagnosis and surgical planning; and *intraoperative* modalities, employed to provide real-time or near-real-time feedback during the procedure itself.

Preoperative Imaging and Surgical Planning

Preoperative imaging provides the basis for patient-specific anatomical understanding and procedural planning (Figure 2.1A). Among the available modalities, CT provides high spatial resolution and excellent visualization of bony and calcified structures, while MRI enables superior soft-tissue characterization and supports advanced functional and diffusion imaging, particularly relevant in neurological and cardiovascular applications.

In the cardiovascular field, imaging protocols capture both anatomical detail and tem-

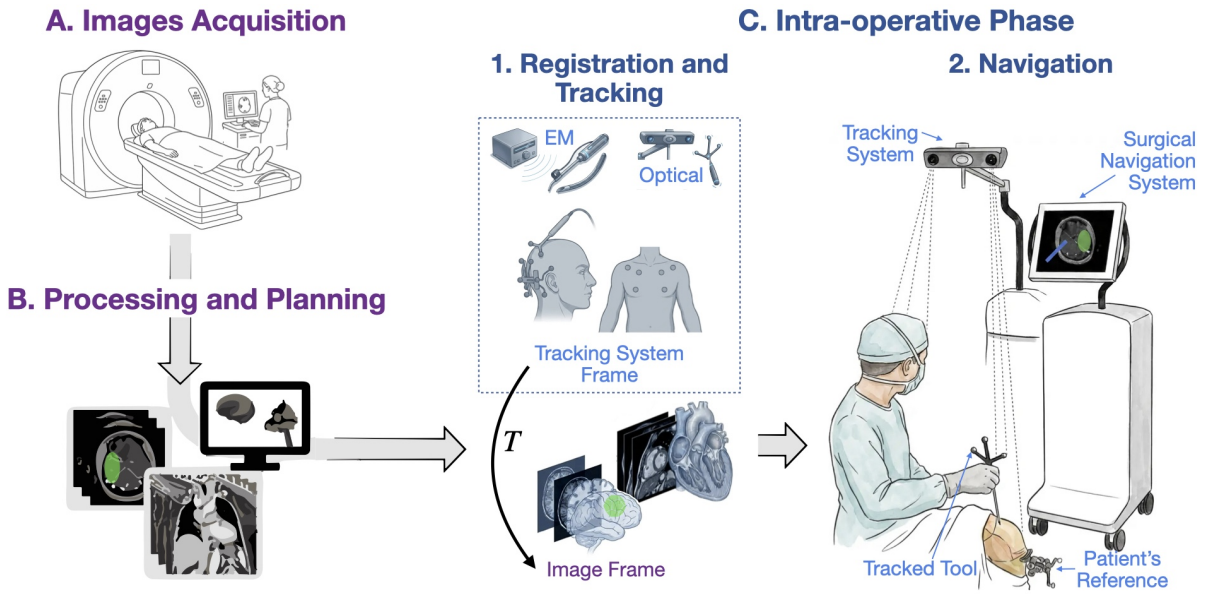


Figure 2.1: Overview of the CAS workflow. The workflow comprises three main phases: **(A)** Images Acquisition: preoperative medical imaging (CT/MRI) of the patient's anatomy; **(B)**Processing and Planning: segmentation and 3D reconstruction of anatomical structures, followed by surgical planning and target identification; **(C)** Intraoperative Phase, which includes: **(1)** Registration and Tracking establishing spatial correspondence between preoperative images and the patient's intraoperative reference frame; and **(2)** Navigation through a two-dimensional (2D) multimodal display that integrates preoperative planning data with intraoperative tool tracking, enabling precise navigation of the patient's anatomy. Adapted and extended from [163].

poral dynamics. Contrast-enhanced Computed Tomography Angiography (CTA) enables visualization of vascular structures, while time-resolved 4-Dimensional Computed Tomography (4DCT) provides information on cyclic cardiac motion. In this setting, 4DCT consists of a sequence of volumetric CT images acquired over a cardiac cycle, typically synchronized with the Electrocardiogram (ECG) (namely ECG-gated 4DCT), allowing the analysis of cardiac motion and cardiac structure deformations.

To support surgical planning, these datasets are processed to extract task-relevant anatomical structures. This is achieved through image segmentation, which transforms raw intensity data into structured representations of organs, vessels, and other regions of interest (Figure 2.1B).

In cardiovascular applications, segmentation commonly targets structures such as the heart chambers, coronary arteries, and major vessels (e.g., aorta and pulmonary arteries), enabling quantitative assessment and procedural planning, such as device sizing or access path definition. Segmentation approaches span from manual delineation to semi-automatic techniques and, increasingly, fully automated learning-based methods. Recent advances in deep learning have enabled the development of robust, large-scale models capable of simultaneously segmenting multiple anatomical structures across different imaging modalities [30, 84, 220]. Such models can be further adapted to specific clinical applications or anatomical targets through fine-tuning on dedicated, in-house datasets.

The resulting segmentation masks are typically converted into 3D surface representations, for example through marching cubes-based reconstruction [122]. These models enable intuitive visualization and form the geometric basis for downstream tasks, including trajectory planning for surgical access, spatial measurements, and image-to-patient registration [185].

Intraoperative Imaging and Surgical Guidance

Intraoperative imaging modalities provide real-time imaging feedback during the procedure itself. In image-guided surgery and PIs, these modalities span from endoscopic video (e.g., laparoscopy, neurosurgery) to X-ray fluoroscopy, ultrasound with a wide range of probes, and Cone Beam Computed Tomography (CBCT), which is based on the same principles as CT but at a lower radiation dose.

For cardiovascular PIs, fluoroscopy and ultrasound are the most commonly used modalities.

Fluoroscopy is most widely adopted for monitoring catheter-based endovascular procedures, intracardiac device deployment and electrophysiological mapping. Fluoroscopy involves the acquisition of a rapid sequence of radiographic images at a defined frame rate. Conventional C-arm systems provide planar X-ray projections with sufficient temporal and spatial resolution for reliable visualization of catheter navigation within the vascular tree or device positioning and deployment during intracardiac procedures. The system offers many degrees of freedom, enabling image acquisition from multiple angles, a fundamental feature for obtaining multiple projections of anatomical structures from optimized and standard perspective views. In the final fluoroscopic image, bones and dense structures appear dark, soft tissues in shades of gray, and metallic objects such as stents, guidewires, and catheters absorb the most radiation and therefore appear as the darkest elements.

However, fluoroscopy presents important limitations. Non-radio-opaque structures, notably soft tissues such as myocardium, vessel walls, valve leaflets, and surrounding organs, exhibit inherently poor visibility, becoming transiently visible only during contrast injection and its subsequent wash-out. Moreover, X-ray projection eliminates depth information: all structures along the beam path are superimposed, making it impossible to discern the three-dimensional spatial arrangement of instruments and anatomical targets. As a result, in numerous precision-guided cardiovascular procedures, even when utilizing conventional fluoroscopy, the level of spatial accuracy remains inadequate and surgical outcomes depend heavily on the operator's experience [133, 146, 185].

Radiation exposure represents a further critical consideration. Fluoroscopy time depends on operator expertise and procedural complexity, and it is managed according to the ALARA (As Low As Reasonably Achievable) principle to minimize dose while maintaining diagnostic adequacy [8]. The cumulative occupational dose for interventional cardiologists represents a recognized health risk, with regulatory frameworks setting annual exposure limits of 20–50 mSv to mitigate long-term stochastic effects [8].

As a result, fluoroscopy is used intermittently rather than continuously, with cumulative imaging time significantly shorter than overall procedural duration. This temporal fragmentation yields only sparse, discrete snapshots of device-anatomy interaction rather than continuous feedback [8].

Ultrasound imaging represents a complementary intraoperative modality with several distinct advantages: portability, absence of ionizing radiation, continuous real-time imaging, and good soft-tissue contrast [198].

In cardiovascular interventions, multiple ultrasound modalities are employed depending on the anatomical target, imaging depth, and procedural requirements. Conventional 2D linear or convex probes are widely used for peripheral vascular access guidance, such as central venous catheterization and arterial line placement, where target vessels are superficial and the probe can be positioned directly over the access site. Transthoracic Echocardiography (TTE) is commonly employed for transthoracic visualization during selected minimally invasive or surgical access procedures, where limited acoustic windows among the ribs permit partial imaging of cardiac structures. For direct cardiac visualization, Transesophageal Echocardiography (TEE) enables improved imaging by advancing a miniaturized ultrasound probe into the esophagus, providing close proximity to the heart and great vessels while avoiding acoustic interference from the chest wall and ribs. The introduction of multiplane and 3D TEE probes has further enhanced interventional guidance by offering volumetric visualization of cardiac anatomy, thereby mitigating the intrinsic limitations of 2D imaging in spatially complex procedures [198]. For interventions requiring direct intracardiac visualization, such as transseptal puncture during catheter ablation or transcatheter valve procedures, Intracardiac Echocardiography (ICE) provides an additional modality: a miniaturized, steerable catheter with an integrated ultrasound transducer is advanced directly into the cardiac chambers, offering high-fidelity visualization of the puncture site and surrounding anatomy [198]. An analogous approach for intravascular navigation is provided by Intravascular Ultrasound (IVUS), which enables cross-sectional imaging of vessel lumen and wall morphology from within the vasculature.

Despite these advantages, cardiovascular ultrasound faces several constraints. Penetration depth of standard examinations is limited to approximately 15 cm, hindering visualization of deep-seated structures, particularly in obese patients [198]. In addition, metallic instruments such as needles and guidewires generate acoustic shadows and reverberation artifacts, degrading image quality and complicating accurate localization of the instrument tip [198]. Furthermore, effective use of ultrasound for interventional guidance requires the simultaneous coordination of probe positioning and instrument manipulation, demanding substantial operator experience and manual dexterity, especially in needle-based procedures where both tasks are performed concurrently.

Real-time semantic segmentation of intraoperative image streams has emerged as a key enabling capability for many downstream tasks. Recent deep learning-based approaches have demonstrated the automatic delineation of relevant cardiac structures in echocardiographic sequences, including the myocardium [226] and anatomical targets such as the valve annulus [154], as well as the detection and tracking of small interventional devices [155]. These methods provide a means to extract higher-level semantic information

from inherently ambiguous imaging data, thereby enhancing situational awareness during intervention.

Taken together, these limitations highlight that no single imaging modality currently satisfies all the requirements for optimal needle or device navigation in cardiovascular PIs [198]. This complementary yet incomplete imaging landscape motivates the integration of multimodal imaging data through registration and fusion frameworks, as well as the augmentation of intraoperative imaging with preoperative anatomical models and real-time tracking information, topics covered in the following sections.

2.1.2. Intraoperative Tool Tracking

Knowledge of the spatial position and orientation, i.e., the *pose*, of a surgical instrument relative to the patient anatomy is a fundamental requirement for any intraoperative navigation system. Without this information, meaningful real-time guidance cannot be provided to the operator.

Tracking, the continuous quantitative estimation of pose over time, can be broadly classified into image-based and sensor-based.

Image-Based Tracking

Image-based, or vision-based, tracking methods estimate the pose of surgical instruments directly from intraoperative image streams, without requiring dedicated hardware integrated into the tool. This represents a practical advantage, as such methods can be implemented as purely software-based solutions on top of existing imaging systems and clinical workflows [172].

For this reason, image-based tracking has been extensively investigated in fluoroscopy-guided cardiovascular interventions, where metallic catheters and guidewires exhibit high radio-opacity and can be reliably identified in X-ray projections. Approaches range from catheter-specific detection exploiting known structural features, such as electrode patterns in electrophysiology catheters [145], to more general representations based on parametric curves [34, 78] or deep learning-based segmentation enabling fully automatic catheter and guidewire detection and centerline extraction from fluoroscopic sequences [10, 233].

Despite significant progress, image-based tracking remains inherently constrained by the projective nature of fluoroscopic imaging; due to the lack of depth information, recovering the 3D catheter pose typically requires either biplane acquisitions or integration with preoperative 3D models, naturally leading to the 2D/3D registration problem. While ef-

fective in static or weakly deforming scenarios, such registration-based approaches must be continuously updated to account for respiratory and cardiac motion in cardiovascular interventions. Despite recent efforts exploring learning-based motion compensation strategies [124], catheter tracking itself is often leveraged as a motion surrogate or registration anchor in these update schemes [9, 29].

Sensor-Based Tracking

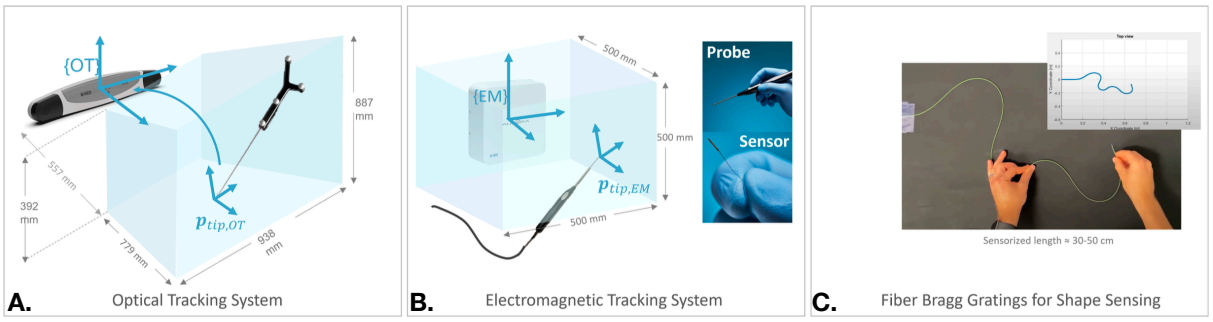


Figure 2.2: Sensor-based tracking system technologies [163]. (A) Optical Tracking System: NDI Vicra (NDI - Northern Digital Inc., ON, Canada) infrared camera tracking a passive marker probe; tip position $p_{tip,OT}$ determined through calibration. (B) Electromagnetic Tracking System: NDI Aurora® field generator tracking an Electro-magnetic (EM) sensor probe; tip position $p_{tip,EM}$ directly estimated. (C) Fiber Bragg Grating Shape Sensing, optical fiber (30–50 cm) with strain sensors; wavelength shifts enable real-time shape reconstruction. Reproduced with permission from Authors [163].

Sensor-based tracking systems provide quantitative pose estimates of surgical instruments independently of the imaging stream and constitute the dominant technology in commercially available surgical navigation systems [192] (Figure 2.2). These systems rely primarily on two sensing modalities: EM tracking, based on electromagnetic field generation, and optical tracking, which employs stereoscopic camera systems (Figure 2.1C.1).

In the context of cardiovascular PIs, these systems enable continuous localization of instruments even in the absence of reliable visual cues from intraoperative imaging modalities.

Optical Tracking Systems (OTSs) operate in the visible or infrared spectrum to estimate the pose of reflective or active markers rigidly attached to surgical instruments. They are widely adopted in CAS applications due to their high accuracy, robustness to environmental noise, and large working volumes, with commercially available systems achieving sub-millimetric accuracy and precision [192]. Their principal limitation is the strict requirement for an unobstructed line-of-sight between the marker constellation and

a fixed camera system in the operating room; marker occlusion during instrument manipulation directly leads to tracking failure. Moreover, OTS implicitly assumes rigid-body behavior: the tracked pose corresponds to the marker-defined rigid body, while any deformation distal to this reference, such as bending of a long and slender catheter, cannot be observed. The working volume depends on the camera system configuration, and typically a surgical tool is tracked at its handle, while its tip position is determined through pivot and spin calibration processes (Figure 2.2A).

EM tracking systems were developed specifically to overcome the line-of-sight limitations of optical tracking. They rely on a field generator producing a spatially varying electromagnetic field of known geometry, which is sensed by miniaturized coils embedded at or near the instrument tip to estimate its pose in real-time. This makes EM tracking particularly well suited to catheter-based and endovascular procedures, where sensors can be placed close to the clinically relevant distal tip and remain observable irrespective of occlusions (Figure 2.2B). The main drawbacks include susceptibility to metallic distortions, limited effective tracking volume, and challenges in integrating field generators into fluoroscopic environments without introducing X-ray artifacts [172]. Recent work addressing these limitations through radiolucent field generator designs has demonstrated substantial reductions in X-ray attenuation while maintaining millimetric tracking accuracy [162].

Optical and EM tracking systems are both well established in surgical domains such as orthopedics and neurosurgery, where the patient anatomy and navigated instruments (e.g., drills, probes, or saw guides) can be reasonably approximated as rigid bodies [105, 177]. Moreover, both tracking technologies can still be applied to needle-based PI and access involving limited deformation [106, 150]. However, transcatheter PIs pose fundamentally greater challenges: long, flexible devices undergo continuous deformation while advancing through a compliant and dynamically moving vascular anatomy, which cannot be fully captured by rigid-body tracking alone, even when distal electromagnetic sensors are employed.

Fiber Optic Shape Sensing (FOSS) represents a third sensor-based modality with qualitatively different capabilities. By embedding multiple strain-sensitive Fiber Bragg Gratings (FBGs) along an optical fiber, FOSS sensing enables reconstruction of the full 3D shape of flexible instruments rather than only a single pose measurement at the tip [172]. This makes FOSS sensing particularly attractive for long, deformable catheters (Figure 2.2C). Hybrid approaches combining FBG shape sensing with EM sensors have been proposed to achieve both highly accurate shape reconstruction and global spatial localization, with reported accuracies on the order of 1 mm [6, 69].

Regardless of the modality employed, the pose of the tracked instrument must ultimately be expressed in a spatial reference frame shared with the anatomical data used for guidance.

2.1.3. Registration Approaches

The preceding subsections have introduced multiple independent coordinate systems involved in modern CAS: preoperative image space (CT, MRI), intraoperative image space (fluoroscopy, ultrasound), and tracking sensor space (OTS, EM, FOSS). For surgical navigation and guidance, these heterogeneous spatial references must be brought into spatial correspondence through *registration*.

Registration is the process of mathematically determining spatial transformations that align data from different sources into a common coordinate system [38]. It is the connective layer between the heterogeneous data streams of a navigation pipeline: preoperative volumetric images, intraoperative imaging, and sensor-derived instrument pose information must all be co-registered to a shared spatial frame of reference allowing the navigation of the patient's body as positioned on the operating bed.

This section first addresses image-to-image registration, enabling image fusion, then examines registration approaches that integrate tracking data for surgical navigation.

Image-to-Image Registration for Multimodal Fusion

Given the fragmentation and complementarity of imaging modalities discussed in Section 2.1.1, cardiovascular PI benefits from the spatial integration of image datasets acquired at different times or using different physical sensing principles.

Image fusion inherently relies on image registration to establish spatial correspondence, whereby one dataset (the *secondary* image) is transformed to align with a *primary* reference image through a sequence of preprocessing and spatial transformation steps.

Once spatial correspondence is established, the co-registered images can be superimposed or co-displayed to provide a unified visualization, often using distinct color mappings or layered representations to preserve modality-specific information [1, 2]. This integration of anatomical and functional information, commonly referred to as *multimodal image fusion* [37], enhances spatial understanding and procedural guidance [168].

The choice of the registration approach depends on both the clinical scenario and the nature of the imaging data. Rigid or affine transformations assume global spatial consistency and are suitable for anatomically rigid regions, whereas deformable (non-rigid) registra-

tion enables spatially varying local transformations to account for tissue deformation and shape changes. In addition, registration strategies differ according to the dimensionality and modality of the data being aligned, such as projection-to-volume, slice-to-volume, video-to-volume, or volume-to-volume registration [118].

In cardiovascular applications, where anatomy is subject to deformation from cardiac motion, respiration, and device-tissue interaction, deformable registration strategies are often required to maintain accurate alignment across modalities [195]. Recent progress in learning-based image registration, particularly deep neural network methods capable of estimating dense and regularized deformation fields, has further improved the robustness and scalability of multimodal image fusion in complex and dynamic clinical settings [37].

From a clinical perspective, the value of fusion-based guidance lies in its ability to overlay preprocedural anatomical or functional information onto live intraoperative imaging. This enables visualization of targets that may not be directly visible in real-time, such as lesions detectable only in arterial-phase CT or MRI, thereby reducing reliance on mental image integration [1, 2]. In cardiovascular interventions specifically, fusion techniques including CT-fluoroscopy [93, 203], MRI-ultrasound [174], and ultrasound-fluoroscopy co-registration [3] have demonstrated clinical benefits such as improved instrument guidance, reduced radiation exposure, and more reliable target localization [198].

However, image fusion alone does not establish the spatial relationship between imaging data and the physical patient or tracked instruments, a requirement addressed through the registration methods discussed in the following section.

Patient-to-Image Registration for Surgical Navigation

In surgical navigation, external sensing devices (e.g., optical cameras, EM field generators, or depth sensors) define the *global reference frame* of the surgical workspace. This corresponds to an *outside-in* tracking paradigm, in which the pose of instruments and the patient are estimated relative to a coordinate frame established by the sensing infrastructure.

Navigation therefore requires explicit calibration and registration of this tracking coordinate system to the patient-specific anatomical coordinate system derived from medical imaging. The instrument pose is first determined in tracking space and subsequently mapped into patient or image space.

This transformation chain's accuracy depends on the intrinsic precision of the tracking modality, the quality of sensor calibration, and the robustness of the registration procedure

itself, making registration a primary determinant of overall targeting precision and clinical reliability.

Surgical registration is most commonly achieved by matching corresponding structures in two spaces. Two broad categories of approach are available: *marker-based* methods, which rely on artificial or anatomical fiducial correspondences, and *markerless* methods, which exploit geometric or photometric surface information acquired intraoperatively [53].

Marker-based Registration Marker-based registration establishes spatial correspondence through a set of identifiable point pairs whose positions are known in both the tracking and the image coordinate systems. Given N such corresponding points, the rigid transformation aligning the two spaces is estimated by minimizing the sum of squared distances between the paired points, yielding a closed-form least-squares solution. The markers themselves may be paste-type skin fiducials [219], surgically implanted bone screws [61], or anatomical landmarks [159] identified interactively with a tracked pointer. Registration accuracy depends on the number, spatial distribution, and localization uncertainty of the fiducials, and increases with the distance of the target from the fiducial constellation. The three marker types present distinct practical trade-offs. Paste-type skin markers are non-invasive but are susceptible to displacement during the procedure, which directly degrades registration accuracy. Implanted markers are more stable and accurate but require a minor surgical intervention for placement. Anatomical landmarks avoid any additional hardware, but their reliable identification requires distinctive surface features, such as the glabella or the orbital margins used in neurosurgery, and their application to soft-tissue anatomies such as the cardiovascular or abdominal regions, where stable and distinctive landmarks are absent, is substantially more difficult.

Markerless Registration Markerless registration methods reduce procedural complexity and avoid invasive marker placement by relying on geometric information acquired intraoperatively, most commonly a surface representation of the patient obtained through a depth sensor or structured-light system, or optically tracked point sampling [53]. This surface is aligned to a surface model extracted from the preoperative image, formulating the problem as a point-set or surface matching task typically solved with iterative algorithms such as Iterative Closest Point (ICP). The accuracy and robustness of surface-based methods remain sensitive to data quality, surface completeness, occlusion, and anatomical deformation, and must be validated using application-specific metrics such as Chamfer Distance or other point-cloud metrics [120].

Recent work has extended the markerless paradigm in two important directions. Prob-

abilistic and global registration strategies improve robustness in anatomies with limited surface features: a fully automatic markerless framework for percutaneous abdominal interventions based on coherent point drift registration of smooth abdominal surfaces has achieved clinically acceptable target localization accuracy without user interaction or fiducials [117]. Simultaneously, markerless surface reconstruction has been integrated directly into the tracking infrastructure, enabling joint reconstruction of patient surfaces and instrument tracking within the same optical coordinate system through synchronized structured-light projection, which eliminates explicit calibration between depth sensors and the tracker and reduces cumulative transformation errors [231].

2.2. Medical Extended Reality (XR)

The CAS methodologies reviewed in Section 2.1 provide the computational infrastructure for surgical navigation: image processing extracts patient-specific anatomy, tracking localizes instruments, and registration establishes spatial correspondence across heterogeneous coordinate systems. However, these capabilities address only the *computational* foundation of surgical guidance. The equally critical challenge of *information delivery*, how guidance is conveyed to the surgeon, remains largely unresolved when confined to external 2D monitors.

Medical XR directly addresses this limitation, serving simultaneously as the display layer and user interface of the broader CAS framework.

2.2.1. Definitions and Technology

The conceptual foundation of XR was laid by Milgram and Kishino [144], who introduced the *Reality-Virtuality Continuum*: a linear spectrum ranging from the purely physical world at one extreme to a fully computer-generated Virtual Reality (VR) environment at the other. Intermediate positions on this continuum correspond to systems that blend real and virtual content in varying proportions. Augmented Reality (AR) occupies the region closest to reality, overlaying digital information onto the user's direct perception of the physical world; Mixed Reality (MR) extends this concept by enabling virtual and real objects to co-exist and interact in the same perceptual space.

The term XR has since been adopted to collectively denote AR, MR, and VR, capturing the entire span of this continuum. This thesis employs XR to refer to this broader spectrum.

Stapleton and colleagues [196, 197] extended Milgram's continuum into a triangular

Mixed-Fantasy Framework by adding imagination as a third axis alongside reality and virtuality. This framework, originally developed for experiential media, carries direct relevance for Medical XR: the goal is not merely to superimpose data onto the operative field, but to shape the surgeon's mental model of a situation that is only partially visible, enabling inference and anticipation beyond what any single sensing modality directly reveals.

This philosophical perspective is explicitly embraced by Navab et al. [157], who observe that since Aristotle's time, humans have relied on all their senses to explore and comprehend their environment, gradually building intuitions from accumulated experiences. A wine glass, for instance, is known to be fragile not because any sense detects fragility directly, but because past observations have been internalized into a tacit understanding. XR thus serves a dual purpose: it brings relevant information to the user's direct attention while enabling the accumulation of sensing-enhanced experiences that deepen environmental understanding and improve performance even without active augmentation [157].

This framing positions XR not as a display technology but as a tool for augmenting cognition, an insight that is particularly consequential in surgical domains where expert performance depends on the integration of sensory cues that are weak, indirect, and partially occluded.

From a hardware perspective, XR systems relevant to surgical guidance can be broadly classified according to their display modality. OST-HMD use transparent optical combiners to superimpose virtual content onto the user's direct view of the real world, preserving natural vision while adding contextual digital information. This approach is exemplified by commercially available devices such as the Hololens 2™ (Microsoft Corporation) or Magic Leap 2™ (Magic Leap, Inc.), which integrate an advanced optical projection system with a suite of onboard sensors including depth cameras, inertial measurement units, and scene-understanding processors. Video See-Through Head-mounted Display (VST-HMD) such as Apple Vision Pro™ (Apple Inc.), by contrast, capture the real-world scene through onboard cameras and composite it with virtual content before presenting the merged view on internal displays; while this sacrifices direct optical transparency, it enables more flexible rendering control, including occlusion handling and photometric consistency between real and virtual elements. Screen-based AR systems represent a third category, consolidating heterogeneous intraoperative information into a unified display without requiring the surgeon to wear a Head-mounted Display (HMD).

2.2.2. Medical Extended Reality Systems

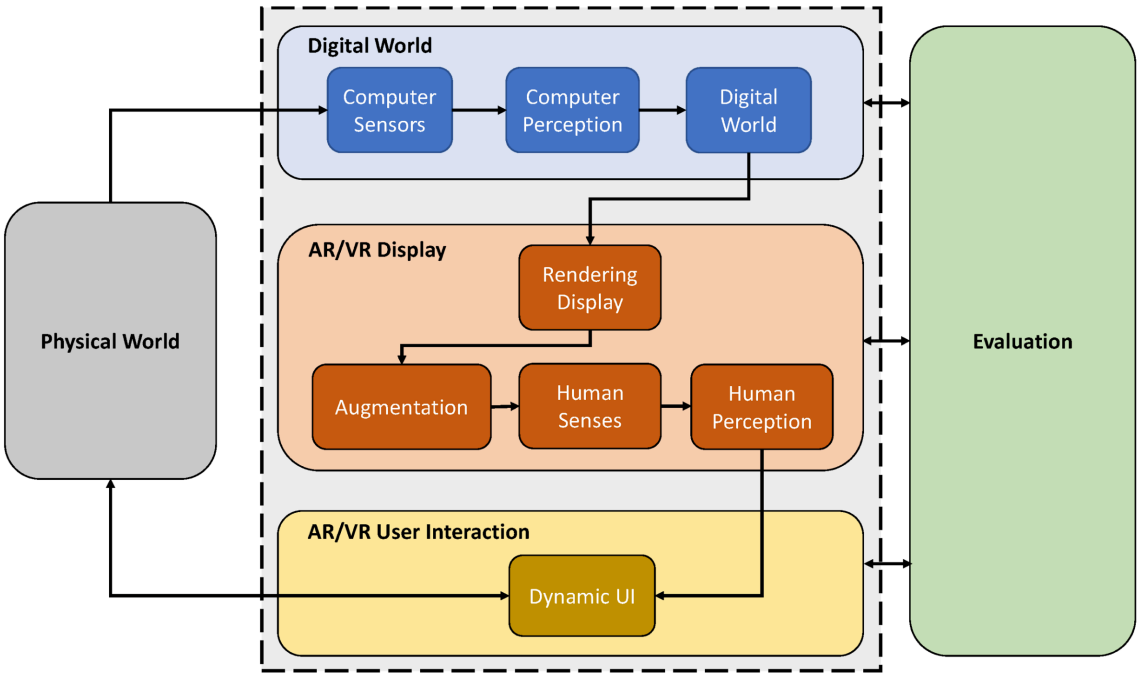


Figure 2.3: The Medical Extended Reality Framework, as outlined by Navab et al. [157], comprises four core components. The system perceives the physical world through its sensing modalities, constructs a *Digital World* representation by processing the collected information, and presents augmented content through the *XR Display* that users perceive and interact with via *XR User Interaction* interfaces. *Evaluation* is essential throughout the system's design and development. Reproduced with permission from Authors [157], published in Journal of Imaging, ©2022 MDPI. Licensed under an open access Creative Commons CC BY 4.0 license.

The systematic development of XR as a discipline applied to medical and surgical domains requires a principled framework that goes beyond display technology to encompass the full pipeline from physical world sensing to user interaction and evaluation. Navab et al. [157] proposed such a framework, the Medical XR Framework, identifying four fundamental components that any medical XR system must address: Digital World, XR Display, XR User Interaction, and Evaluation.

While the physical environment itself is not a component of the XR system, the system must sense and interpret it to generate meaningful augmentation. *The quality of augmentation fundamentally depends on the system's ability to understand and model the physical world*: the more accurately the system can perceive, represent, and semantically interpret the environment, the more effective and contextually relevant the augmentation

becomes [157].

The *Digital World* component concerns the computational representation of the physical environment through digital models, often referred to as digital twins, of relevant entities such as patient anatomy and surgical instruments. This representation goes beyond geometric replication to encompass semantic understanding of the scene: identifying anatomical structures, recognizing surgical instruments, understanding their spatial relationships, and interpreting ongoing interactions. This component maps directly onto the CAS foundations reviewed in Section 2.1: image processing, tracking, and registration all contribute to building and maintaining this digital representation.

The *XR Display* component addresses how computed digital information is transformed into user-comprehensible augmentation through visual, auditory, and haptic feedback. A defining feature of this component is its emphasis on perceptual realism, ensuring that virtual content feels natural yet remains distinguishable within the physical context. In the visual domain, techniques such as adaptive transparency and depth cueing help maintain clear distinction between virtual and real elements, minimizing perceptual confusion in high-risk environments. Crucially, the XR Display is inherently multisensory: haptic feedback can convey force or contact information, while auditory cues can enhance spatial localization of instruments or anatomical targets [157], modality that will be further developed in Section 2.4.

The *XR User Interaction* component addresses the complexities of engaging with digital systems in sterile, high-stakes environments such as OR, where traditional input devices are often inadequate. The framework emphasizes adaptive, context-aware interfaces that align with surgical workflow, employing hands-free interaction methods such as gesture recognition, gaze tracking, and voice commands.

The *Evaluation* component encompasses rigorous assessment of system functionality, accuracy, usability, and safety throughout design, development, and clinical translation. Evaluation combines subjective measures, such as workload assessments and user satisfaction surveys, with objective metrics including latency, precision, and reliability. Iterative validation with end-users is essential, as system designers may otherwise develop solutions that contradict the intuitions of the clinical experts they aim to support.

These four components collectively define the design space of medical XR system and provide a structured methodology for developing, comparing, and validating systems across clinical applications. Importantly, they reinforce that XR is not merely a visualization layer but an integrated system that must be co-designed with its CAS computational backend and validated end-to-end with medical experts.

2.2.3. Registration for XR-Guided Surgery

XR systems introduce an additional spatial relationship beyond those described in Section 2.1.3: the transformation between the display coordinate system and the physical world. For OST-HMDs, this involves calibrating the virtual rendering environment to the device's internal tracking frame; for projector-based systems, determining the projector's pose relative to the surgical scene; and for video-based AR, camera-projector calibration and real-time pose estimation [125].

Current HMD devices use onboard depth cameras and inertial sensors to estimate device pose within the environment, relying on Simultaneous Localization and Mapping (SLAM), a computational framework that simultaneously constructs a spatial map and localizes the device within it [67, 70]. SLAM maintains stable anchor points for virtual content, ensuring augmented elements remain spatially consistent as the user moves. However, SLAM alone does not establish the spatial correspondence required for surgical navigation, which necessitates an additional calibration chain relating patient anatomy, preoperative imaging, and the HMD's coordinate frame to intraoperative imaging or external tracking systems [70].

Onboard sensors enable *inside-out* tracking that fundamentally alters the registration workflow presented in Section 2.1.3 (*outside-in* tracking). While external systems establish a stable global reference frame by tracking both patient and instruments, *inside-out* tracking paradigm define a device-centric frame through SLAM, which change when the user changes position within the environment [67, 70].

Marker-based approaches adapt to inside-out tracking through multimodal fiducials visible to both imaging modalities and HMD cameras, enabling registration without external infrastructure [70, 90]. **Markerless** registration exploits HMD depth sensors to capture patient surface geometry, though with lower accuracy than dedicated scanners due to sensor and ergonomic limitations [125].

Solving this calibration chain enables *in-situ* visualization of anatomical models, navigational cues, or virtual screens with ultrasound data [216] directly in the surgeon's view without line-of-sight constraints.

While inside-out tracking enables more flexible use in the OR, it introduces distinct challenges. Registration must account for SLAM localization uncertainty and temporal drift, absent in external tracking systems [70, 125]. Errors accumulate over time and space, demanding careful correction for millimeter-scale guidance and making accuracy dependent on patient motion, anatomical deformation, HMD drift, and display calibration.

These challenges motivate hybrid registration strategies combining multiple modalities, such as laser-based [170] or external depth cameras [163], with continuous registration refinement in time to maintain spatial correspondence.

2.2.4. Promising Clinical Applications

Medical XR has demonstrated substantial clinical benefits across surgical domains, with several applications maturing from phantom studies to prospective clinical evaluations and, in selected cases, regulatory-approved commercial products [130].

The clinical deployment of XR-based surgical navigation has accelerated markedly in recent years, expanding across diverse clinical domains, particularly in robotic and semi-autonomous surgical systems. Current applications span neurosurgery, cardiothoracic surgery, orthopedics, gastrointestinal interventions, and ear-nose-throat procedures, reflecting both the maturity of rigid-anatomy navigation and the expanding scope toward deformable-tissue settings [186].

Preoperative planning and education Medical XR has been extensively validated for preoperative visualization of patient-specific 3D anatomical models, offering improved spatial understanding as compared to conventional 2D displays and demonstrating significant educational value [151, 229]. In cardiovascular applications, platforms integrating cardiac CT and echocardiography data into collaborative AR environments support treatment planning and procedural rehearsal [94, 169], with cross-reality prototypes enabling seamless transitions across the Reality-Virtuality Continuum demonstrating sustained use in pilot studies [4].

Intraoperative guidance in rigid anatomies The most thoroughly investigated and clinically impactful application of medical XR as a surgical navigation system involves procedures targeting rigid or quasi-rigid anatomical structures, where the spatial relationship between the preoperative model and the intraoperative anatomy remains stable throughout the intervention. Neurosurgery and orthopedics represent the primary domains due to their amenability to accurate registration. Marker-based XR neuronavigation using optical tracking cameras has achieved targeting errors between 0.5 and 3.5 mm in controlled settings through successful fiducial marker placement for preoperative registration [186]. Recent work on external ventricular drain placement demonstrated a fast hybrid registration workflow combining landmark alignment with markerless surface tracing, achieving total registration time below two minutes with high usability scores [56].

The choice of visualization strategy significantly impacts navigational accuracy. In the ex-

ternal ventricular drain study, dynamic tool-tracked visualization markedly outperformed static in-situ overlays, reducing median target-to-tip error from 17.5 mm to 4.3 mm and angular deviation from 4.3° to 2.3° [56]. These results suggest that dynamic instrument-aware visualization following rapid registration may satisfy the demands of emergency neurosurgical navigation. In neurosurgery, the combination of automated CT-based segmentation, preoperative planning, and XR visualization through an OST-HMD improved mean targeting accuracy of external ventricular drainage by 42% (9.6 mm vs. 16.2 mm) as compared to the conventional blind technique, while maintaining computation times below three minutes [164]. In spinal surgery, CT-pre-planned pedicle screw trajectories visualized through Hololens 2^{TM} guidance achieved a mean angular error of 3.38° ± 1.73° and an insertion-point error of 2.77 ± 1.46 mm on vertebral phantoms [119]; a subsequent cadaver study found no significant difference in accuracy between XR-navigated drilling and gold-standard fluoroscopy, supporting its potential as a radiation-free alternative in selected procedures [153].

Notably, the benefits of XR-based navigation are not limited to HMD. Screen-based AR systems integrating OTS-based tool tracking have also been shown to substantially outperform conventional fluoroscopy-guided freehand techniques in sacral nerve stimulation. Phantom studies reported significant reductions in procedure time, needle punctures, and complete elimination of intraprocedural X-ray usage, highlighting the potential of AR navigation, independent of the display modality, to improve efficiency and reduce radiation exposure for both patients and operators [150].

2.2.5. Intraoperative Guidance in Cardiovascular Procedures

Unlike rigid anatomies where XR guidance has demonstrated consistent clinical benefits, cardiovascular applications face fundamental challenges that have limited their translation to intraoperative use. While XR technologies in vascular surgery show promise for preprocedural planning, education, and telementoring, real-time guidance remains constrained by registration drift, virtual cue instability, and the absence of validated model-updating strategies suitable for continuously deforming anatomy [39, 112]. Current implementations rely on static preoperative anatomical models that progressively diverge from intraoperative reality as tissue deforms.

Several research groups have investigated XR guidance in cardiovascular procedures, revealing these fundamental limitations. In endovascular aortic repair, EM tracking integrated with Hololens 2^{TM} enabled radiation-reduced navigation by overlaying preoperative CT anatomical models and virtual angioscopy views onto the operative field,

with the system tracking catheter tip position to eliminate reliance on fluoroscopy for device localization [58]. Similarly, AR guidance for transcatheter structural heart interventions achieved initial automated registration based on spine landmarks visible in both CT and fluoroscopy, with sub-millimetric accuracy and computational times compatible with intraoperative use, providing real-time 3D visualization of catheter position to address depth perception limitations inherent to 2D fluoroscopic guidance [121]. However, both approaches remained fundamentally constrained by rigid registration between preoperative data and the patient, as vessel deformations induced by catheter navigation, respiratory motion, and tissue displacement were not accounted for in the static virtual overlays.

Comparative studies have highlighted perceptual benefits of 3D XR visualization over conventional 2D monitor displays for endovascular catheter navigation [225]. When controlling teleoperated robotic catheters, operators using Hololens 2^{TM} for 3D feedback achieved lower tracking errors, reduced targeting errors, and shorter procedure durations compared to standard monitor-based visualization, suggesting that depth perception provided by XR displays enhances spatial awareness during catheter manipulation. However, these advantages were demonstrated exclusively in static vascular phantoms where preoperative models remained valid throughout the navigation task.

This limitation is compounded by the incomplete spatial information provided by standard catheterization imaging (fluoroscopy and ultrasound), which is insufficient to continuously re-register or update 3D anatomical representations during intervention. Model-updating strategies to address this challenge are reviewed in Section 2.3.

Consequently, while XR guidance in rigid anatomies achieves tracking and target-reaching accuracies of 1 to 4 mm with angular deviations between 1° and 3° [12], clinically acceptable for many procedures, extending this capability to deformable cardiovascular anatomies requires continuous model updating strategies capable of operating under varying degrees of intraoperative observational data, approaches reviewed in the following section.

2.3. Anatomical Updating Strategies for Surgical Navigation

Preoperative anatomical models provide the foundation for surgical planning, but deformations during intervention progressively invalidate these static representations. Soft-tissue displacement from instrument contact, respiratory motion, and cyclic cardiac con-

traction introduce variations that are not always captured in conventional preoperative imaging.

Cardiac wall displacement during the cardiac cycle has been quantified across multiple studies, with motion magnitudes typically less than 5 mm but averaging 1–2 mm across different cardiac chambers and phases [76, 161, 210]. These displacements vary by anatomical location: the left ventricle exhibits medio-lateral shifts of 0.5–0.8 mm and superior-inferior motion up to 2.7–3.1 mm, while pericardial boundaries show directional displacements of 0.4–0.8 mm [76, 161, 210]. Similarly, vessel deformations induced by catheter and guidewire insertion during endovascular procedures have been measured at 2–5 mm on average, with patient-specific variability reaching up to 9 mm in tortuous anatomies [64].

When these deformations are not accounted for, XR overlays diverge from surgical reality, undermining the spatial and temporal correspondence required for accurate guidance. The challenge of maintaining correspondence between preoperative anatomical models and the evolving intraoperative anatomy, identified as the central technical barrier to reliable XR-guided surgery [72], has motivated the development of deformation modeling approaches specifically for surgical navigation.

As outlined in Section 2.2.2, the modeling of the physical world into a digital representation plays a central role in medical XR system: the quality of augmentation is fundamentally bounded by the accuracy of the underlying world model.

When intraoperative observations are available and sufficiently informative, preoperative model updating can be achieved through image-based approaches..

However, a more fundamental problem arises when the data required to update the model is simply not observable through available sensing modalities. For instance, in fluoroscopy-guided endovascular navigation conducted without contrast agent injection, vessel deformations induced by catheter contact remain entirely invisible: the catheter itself is radio-opaque and clearly visualized, but the surrounding vascular anatomy does not produce sufficient contrast for geometric reconstruction. In such scenarios, no amount of algorithmic sophistication in image processing can recover information that was never captured by the imaging modality in the first place. This observability gap motivates the exploration of model-driven approaches that can infer or predict anatomical states based on partial observations, physical principles, or learned priors.

The following subsections review image-based and model-driven strategies for addressing this challenge.

2.3.1. Image-Based Approaches

When intraoperative imaging modalities provide direct visualization of anatomical structures, model updating can be formulated as a multimodal image registration problem (Section 2.1.3).

Ultrasound has emerged as a particularly attractive modality for this purpose due to its real-time acquisition capability, absence of ionizing radiation, and ability to visualize soft tissue structures. Calibrated ultrasound imaging has been successfully integrated into complete intraoperative updating pipelines for minimally invasive liver surgery, where particle filter-based calibration establishes the spatial transformation between the ultrasound image plane and the surgical coordinate frame [36]. Intraoperative anatomical landmarks identified in ultrasound images drive two-level surface warping algorithms that update preoperative models while preserving anatomical plausibility, achieving clinically acceptable targeting accuracies below 2 mm. Coarse-to-fine hierarchical registration strategies further improve robustness to large anatomical shifts common in abdominal procedures [211].

In laparoscopic settings, learning-based frameworks can estimate temporal deformation fields directly from endoscopic video sequences, leveraging temporal coherence to infer physically plausible deformation trajectories [217]. Graph Neural Network approaches extend this capability by predicting the deformed states of occluded or temporarily invisible structures based on observed motion of connected visible regions [48], offering novel methodologies for dynamic environments where continuous observation is not guaranteed.

Despite these advances, image-based updating approaches share a fundamental dependency: they require that the relevant anatomical information be visible and extractable from the available imaging modality. In contexts where key structures remain unobservable, such as vessel deformations in non-contrast fluoroscopy, or deep anatomical layers occluded by superficial tissue, image-based methods alone cannot provide anatomical model update. This observability constraint directly motivates the second category of approaches, which rely on physics-based simulation to infer unobservable states from partial data and known mechanical principles.

2.3.2. Interactive Physics-based Modeling

Biomechanical simulation has achieved substantial progress in modeling cardiovascular tissue mechanics, contact interactions, and dynamic anatomical behavior. Patient-specific computational approaches can predict wall shear stress in aneurysmal aortas [165], simulate transcatheter valve deployment through coupled fluid-structure interaction [123], and

compute arterial deformations during guidewire insertion [64].

These methods have demonstrated clinical validity when validated against intraoperative ground truth data [65, 123, 165], achieving errors as low as 2.1 mm in predicting guidewire positions [64]. However, these accurate predictions come at prohibitive computational costs: finite element simulations of vascular interventions require 5.5 to 7 hours per case [64, 123], rendering them incompatible with intraoperative decision-making. Consequently, these capabilities have remained largely confined to offline analysis, preoperative planning, and post-hoc validation.

The challenge lies in bringing biomechanical modeling into the intraoperative loop, creating interactive simulations that update in real-time and provide surgeons with dynamic, physically grounded representations of the evolving surgical situation. This integration is particularly critical for cardiovascular interventions, where static anatomical representations and observability gaps in current XR-guided navigation systems (Section 2.2.4) directly motivate the need for real-time, physics-based model updating.

Biomechanical Modeling Foundations Biomechanical modeling enables stress and displacement analysis of deformable bodies. In the context of the present work, it is employed to compute the mechanical response of the vascular wall to catheter advancement within the vascular lumen. This approach requires the numerical solution of differential equations governing compatibility and equilibrium, complemented by constitutive laws that describe material behavior. The solution is obtained based on initial conditions and Boundary Conditions (BCs), where BCs are derived from intraoperative observations [176] and can be imposed either as prescribed displacements or prescribed loads.

The Finite Element Model (FEM) is the most widely adopted approach due to its effectiveness in handling complex geometries and material behaviors. When adequately constrained and mechanically informed, FEM-based models can reproduce the viscoelastic response of hepatic, prostatic, renal, and neural tissues with clinically relevant accuracy [23, 99, 209]. Application in real surgical settings encounters two fundamental difficulties. First, acquiring BCs during surgery remains challenging: kinematic BCs require establishing point-to-point correspondences between the preoperative anatomical model and the deformed intraoperative tissue surface in order to impose known displacements at identified locations, while loading BCs necessitate precise measurement of tool-tissue interaction forces. Second, directly solving FEM-based biomechanical models presents computational limitations, with solution times often incompatible with real-time requirements [72].

Real-time Simulation Strategies In traditional FEM-based approaches, accuracy in the quantification of the stress and strain fields is key, whereas the computational expense of simulations is secondary. Implicit time integration schemes, iterative nonlinear solvers, and fine mesh discretizations achieve converged solutions for offline engineering applications. Intraoperative surgical navigation, however, imposes fundamentally different requirements: simulations must run at interactive rates (25-60 Hz for visual feedback, or 1 kHz for haptic rendering), remain stable under user interaction and contact events, and integrate heterogeneous models, sensors, and visualization components within a unified pipeline. Three distinct strategies address these requirements:

Mass Spring Model (MSM) offer a computationally efficient alternative by approximating continuum mechanics through networks of concentrated masses connected by springs [75]. Extending Hooke's law to each mass point avoids the element-wise preprocessing required by FEM and enables near real-time simulation of large deformations. However, MSM discretization introduces fundamental limitations: the voids between masses and springs contain no material representation, precluding the continuous displacement fields and compatibility conditions that govern FEM formulations. This structural simplification requires specific optimization methods to approximate realistic global mechanical behavior [228]. Despite these limitations, MSMs remain advantageous for tasks involving topology modifications, where cutting, grasping, or suturing can be implemented by simply removing or adding connections between vertices [16, 230]. Consequently, MSMs have been primarily adopted in medical training simulators, where computational efficiency outweighs physical accuracy [108].

Data-Driven Surrogates Machine learning approaches accelerate physics-based simulation by approximating the input-output mapping of FEM solvers at a fraction of the computational cost. Neural network surrogates trained on datasets spanning various BCs, material parameters, and geometries produce deformation predictions in milliseconds, enabling interactive performance for problems that would otherwise require seconds or minutes per timestep [73, 142]. In the cardiovascular domain, purely data-driven surrogates face fundamental limitations due to training data scarcity. Unlike abdominal organs visualized through repeated laparoscopic procedures, dynamic cardiovascular anatomy during catheter-based intervention is rarely captured at sufficient spatiotemporal resolution. Physics-informed neural networks that embed conservation laws and constitutive equations aim to improve generalization and data efficiency [73], though their robustness in highly dynamic scenarios remains an active research area.

Optimized FEM solvers and asynchronous multithreading enable interactivity while maintaining FEM accuracy, requiring rethinking of both numerical methods and the con-

ceptual framework for tissue mechanics, device-tissue contact, and user interaction [50]. Constraint-based formulations, in which contacts, attachments, and user-applied forces are modeled as mechanical constraints solved simultaneously within a unified complementarity framework, enable robust handling of complex surgical interactions including grasping, pulling, cutting, and catheter navigation [167]. Open-source software frameworks have enabled practical realization of this approach.

Simulation Open Framework Architecture (SOFA) [54] has become the standard platform for interactive medical simulation research. Its scenegraph-based, multi-model architecture decomposes each simulated object into independent components, degrees of freedom, internal forces, collision geometry, visual rendering, and haptic interaction, allowing multiple representations of the same physical object to coexist and be optimized independently while maintaining consistency through runtime mapping mechanisms. Transparent CPU and GPU implementations enable computationally intensive operations to be offloaded to graphics hardware without changes to high-level simulation logic [41]. Contact modeling strategies for real-time FEM-based surgical simulations, including penalty forces, Lagrange multiplier constraints, and prescribed surface displacements, offer different trade-offs between computational speed, numerical stability, and accuracy [204].

Parameter Estimation and Model Updating Environments with observability-gap or data scarcity hinder estimation of BCs and patient-specific mechanical parameters needed to build high-fidelity simulations. The constraint-based and image-driven updating strategies reviewed assume availability of intraoperative observations to correct or refine simulation state. In their absence, simulations must rely entirely on internal mechanical models and assumed parameters. Bayesian optimization frameworks address this challenge by estimating unknown BCs from partial surface observations, identifying patient-specific biomechanical models that match observed intraoperative deformations [143, 205]. Inverse estimation techniques can further refine tissue mechanical properties from observed deformations [87], closing the loop between observation, parameter identification, and predictive simulation.

Despite these advances, cardiovascular interventions present a distinct challenge: maintaining spatial awareness of highly deforming, cyclically moving anatomy that is only partially observable through projection imaging.

2.4. Medical Sonification

The preceding sections reviewed physics-based modeling strategies that address the spatial requirements for XR-based surgical navigation in the cardiovascular domain. This section addresses the second methodological aim: the design of **multisensory displays** that meet the temporal requirements outlined in the Introduction (Chapter 1).

Effective XR guidance requires translating computational outputs into representations that surgeons can interpret and act upon in real-time. While visual augmentation has been the primary focus, the auditory channel offers complementary advantages for conveying temporal and dynamic information.

Sonification, the process of conveying information through sound, has demonstrated value in surgical tasks including situational awareness [137], tool alignment [20, 138], and anatomical analysis [139]. This section reviews the foundations of medical sonification and establishes how the methodologies developed in this thesis extend its application to cardiovascular navigation.

2.4.1. Visual Channel Limitations and Multisensory Perception

The human perceptual system is inherently multisensory. Vision, audition, touch, and proprioception operate in parallel, each contributing unique information that is integrated at multiple neural levels to form a coherent representation of the world [52]. Multisensory integration, particularly including auditory cues, enhances performance across a wide range of tasks through attentional mechanisms and modulatory effects on working memory [98, 199]. Delivering information through multiple sensory channels, rather than overloading a single one, leads to better performance [173].

Current surgical navigation systems overwhelmingly prioritize visual display. However, the visual channel in minimally invasive surgery and PIs is already saturated: endoscopic or fluoroscopic images, instrument visualization, patient monitors, preoperative plans, and XR overlays all compete for attention. Adding more visual information increases clutter, degrades attention, elevates cognitive load [46], and can result in *inattentional blindness*, where critical information goes unnoticed despite being visually present [88, 127].

The auditory system offers a complementary channel with distinct advantages for dynamic, time-varying information. Unlike vision, which requires directed attention and is sensitive to occlusion and field-of-view constraints, audition is omnidirectional and operates continuously. The temporal resolution of hearing far exceeds that of vision, enabling detection of rapid changes and subtle temporal patterns [27]. Sound can encode multidi-

mensional data through frequency, amplitude, and spectral patterns, which the auditory system perceives with high precision as pitch, rhythm, and timbre. Its omnidirectionality enables perception independent of head orientation, supporting continuous real-time interaction.

Auditory perception is particularly well-suited to monitoring tasks, event alerting, and conveying rhythmic or periodic information, precisely the characteristics of many surgical navigation scenarios. These advantages have increasingly motivated the use of auditory displays for presenting navigation data and tool-tissue interactions [135], with recent studies demonstrating effectiveness when integrating auditory with visual cues [107, 182].

Medical sonification research can be broadly categorized into two methodological approaches: parameter-mapping sonification and physics-driven model-based approaches.

2.4.2. Parameter-Mapping Approaches

Parameter-mapping Sonification (PMSon) represents the most straightforward approach to auditory display, establishing direct mappings between data features and sound synthesis parameters. In surgical navigation contexts, PMSon has been applied to convey distances to target structures, proximity to risk zones, and alignment states, with the majority of approaches relying on proximity-based mappings where distance modulates pitch, amplitude, or pulsing rate [20].

Methods for transforming basic data inputs, such as multistate signals [136] or spatial mapping [138, 234], have been explored and proven reliable for straightforward navigation scenarios. For instance, in pedicle screw placement, Matinfar et al. [138] demonstrated that four degrees of freedom could be encoded through frequency modulation synthesis: two translational coordinates defining the entry point position on the planned trajectory's perpendicular plane, and two angular deviations describing tool orientation relative to the planned trajectory. This approach achieved accuracy comparable to visual navigation while reducing the surgeon's need to focus on external displays.

However, these approaches struggle to represent complex data such as anatomical geometry or biomechanical tissue properties. PMSon is limited by the perceptual interdependence of sound attributes, nonlinear relationships between sound parameters and perception, and the restricted number of distinguishable sound dimensions available to encode high-dimensional surgical data [82]. Moreover, conventional parameter-mapping approaches, where distances map to beep frequencies, proximity to color changes, or warning states to discrete alarms, operate at a level of abstraction divorced from physical intuition. They require surgeons to learn arbitrary associations, imposing cognitive

overhead and reducing naturalness of interaction [81]. In high-stakes surgical contexts, where attention must be focused on the primary task and response times are critical, such abstractions represent a barrier to effective human-machine collaboration.

2.4.3. Physics-Driven Model-Based Approaches

Human perception is fundamentally shaped by interaction with the physical world. We intuitively understand how objects respond to forces, how materials deform, how structures resonate, and how motion unfolds over time. This physical intuition is deeply embedded in our sensorimotor system, enabling us to predict outcomes and respond to events with remarkable speed [60].

Research in ecological acoustics demonstrates that humans are exquisitely sensitive to invariant patterns in sensory input corresponding to physical regularities [60]. When feedback is grounded in physical laws, it taps into these deep perceptual sensitivities, enabling faster learning, more robust performance, and more confident decision-making. The human auditory system processes spectro-temporal features within fractions of a second [147, 208], and sounds closely mirroring real-world auditory experiences are perceived as more intuitive, with effectiveness depending on the relationship between sound, gesture, and users' prior experiences with everyday objects [35, 116, 201].

Physics-based sound synthesis, generating audio by simulating vibration of virtual objects such as membranes, strings, or resonators, provides a principled framework that addresses the limitations of arbitrary parameter mapping. By embedding surgical data within parameters of a physical sound model (e.g., membrane tension, damping coefficients, excitation patterns), the resulting audio naturally exhibits temporal, spectral, and dynamic characteristics associated with real-world sound-producing phenomena. Listeners do not need to learn arbitrary mappings; the sounds are inherently meaningful because they conform to the physics of vibration and acoustics that humans have internalized through everyday experience [135].

Integrating physics-based sonification with visual XR creates a truly multimodal guidance system in which each perceptual channel contributes information aligned with its strengths: vision for spatial layout and static anatomical context, audition for temporal dynamics and interaction state [47]. This distribution of cognitive load is not merely additive; it is synergistic, leveraging the brain's capacity for multisensory integration to enhance overall situational awareness and decision-making performance [199].

Model-Based Sonification To address the limitations of PMSon, Model-based Sonification (MBS) offers a more sophisticated framework. Rather than prescriptively mapping data to sound parameters, MBS transforms data into sound through dynamic mathematical models that respond to user interaction [81, 82]. The data becomes embedded within a virtual physical system, a *"sound-capable object"*, whose dynamic behavior generates acoustic output when excited by user actions.

While sonification of high-dimensional medical imaging has been explored [66, 81], usability challenges often limit accessibility, particularly for non-musical users.

From Tissue to Sound Framework A recent paradigm shift introduced by Matinfar et al. [140] addresses usability challenges through the *"From Tissue to Sound"* framework, which establishes a comprehensive methodology for mapping tissue characteristics extracted from medical imaging to auditory displays. This approach employs a physics-based sound model composed of a network of multiple oscillators, whose mechanical properties, such as damping, stiffness, and mass, are defined and mapped from tissue characteristics extracted from imaging data.

The framework operates by translating the physical characteristics of anatomical tissues via medical imaging into a physics-driven sound model, utilizing an interconnected network of oscillators with multiple geometrical structures. When users interact with this model through excitation gestures, the system generates acoustic profiles that reflect underlying tissue properties: stiffer tissues produce distinct sounds compared to softer ones, dense regions sound different from sparse areas, and structured tissues yield different acoustic responses than scattered distributions [140].

The *"From Tissue to Sound"* paradigm provides several advantages over conventional PMSon approaches. First, it enables representation of high-dimensional data without arbitrary dimensional reduction, as the oscillator network naturally accommodates complex spatial and material distributions. Second, the physical basis of the sound model ensures perceptual coherence: the acoustic response emerges from simulated mechanical interactions rather than ad-hoc parameter assignments. Third, the interactive nature of the model supports exploratory data analysis, allowing users to actively probe anatomical structures through gestural excitation and receive immediate acoustic feedback that reflects local tissue characteristics.

This MBS approach has been validated in close collaboration with medical experts across realistic surgical task scenarios, including retinal interventions [139] and brain tumor localization [182], demonstrating the effectiveness of auditory feedback in enhancing intra-

operative perception, decision-making, and performance during highly sensitive surgical tasks.

Despite these advances, a fundamental gap remains: current sonification methods operate on static anatomical data and cannot represent tool navigation in dynamically deforming tissues during PIs. Existing frameworks, especially *"From Tissue to Sound"*, rely on preoperative models and lack integration with real-time tissue deformation.

3 | Motivation and Fundamental Impact

3.1. Main Motivation

The preceding chapter has established the state of the art in CAS, XR-based surgical navigation, tissue deformation modeling, and medical sonification. While substantial progress has been achieved in each domain independently, their integration into cardiovascular PIs reveals fundamental challenges that distinguish this application from established use cases in other fields.

The gap between preoperative and intraoperative scenarios, identified as the first limitation for intraoperative reliability of XR systems in endovascular procedures (Section 2.2.5), is not merely a matter of visualization fidelity or spatial accuracy; it represents a fundamental failure of current medical XR system to capture what is truly central to intraoperative navigation: the *surgical interaction* itself.

For example, when a catheter navigates through a tortuous vessel, the interaction reveals information that transcends static anatomical geometry. Vessel wall deformation under catheter contact encodes not only the catheter's current position, but the mechanical compliance of the tissue, distinguishing calcified from healthy vessel segments. Frictional resistance during catheter advancement varies with vessel wall composition, surface moisture, and navigation velocity, revealing textural properties invisible to imaging. When a guidewire buckles against vessel bifurcation, the buckling pattern and resistance encode the angle of approach and local vessel stiffness.

Similarly, when a needle approaches the beating heart, the temporal variation in proximity to the myocardium during the cardiac cycle determines not merely distance, but the optimal phase for safe advancement. These are not properties of the static anatomy; they are properties of the *interaction* between instrument and tissue, a dynamic cascade of causal events governed by physical laws.

Current XR systems focus on modeling anatomy, updating its spatial representation when observational data from intraoperative imaging systems permits. What remains unmodeled is the interaction event itself: the mutual physical process by which the surgical tool perturbs the anatomical state and the anatomical state constrains the tool's motion, encoding compliance, friction, and dynamic relationships within the complex, variable surgical domain that anatomical models alone cannot capture.

3.2. Surgical Interaction Modeling

The fundamental impact of this thesis is to introduce an additional layer within the Medical XR Framework for surgical navigation system: the explicit modeling of surgical interaction (Figure 3.1). Physics-based modeling of this interaction produces two distinct types of information that serve complementary roles within the Medical XR Framework (Section 2.2.2): anatomical updates that refine the digital representation of the physical world, and interaction features that drive perceptual delivery through display modalities.

First, modeling surgical interaction addresses the spatial challenge in the presence of observational constraints from intraoperative imaging. The observability gap inherent to fluoroscopy-guided transcatheter PIs, where instruments are visible but vessel deformations remain invisible without contrast agents, prevents purely image-based model updating. However, the interaction itself is partially observable, as catheter shape and contact forces can be reconstructed from other sensing modalities. Physics-based simulation of the interaction event uses these partial observations to predict unobservable tissue states, substituting missing imaging data with physically grounded prediction. These updates feed into the Digital World representation, maintaining spatial correspondence when direct observation is incomplete.

Second, modeling surgical interaction enables perceptual representation of guidance information through physics-driven sonification under strict temporal requirements. Cardiovascular procedures involve anatomical changes at frequencies that challenge visual perception, while the visual channel is already saturated. The *"From Tissue to Sound"* paradigm introduced physics-based sonification for static tissue [140], but has not been extended to dynamic intraoperative scenarios. The causality inherent to surgical interaction provides a natural foundation for this extension. When the same biomechanical model that simulates the interaction also drives acoustic synthesis, the resulting sound naturally reflects interaction dynamics: proximity to moving structures, contact events, tissue resistance, frictional changes. This delivers high-temporal-resolution guidance through the auditory channel with high intuitiveness, precisely because the sound emerges from the

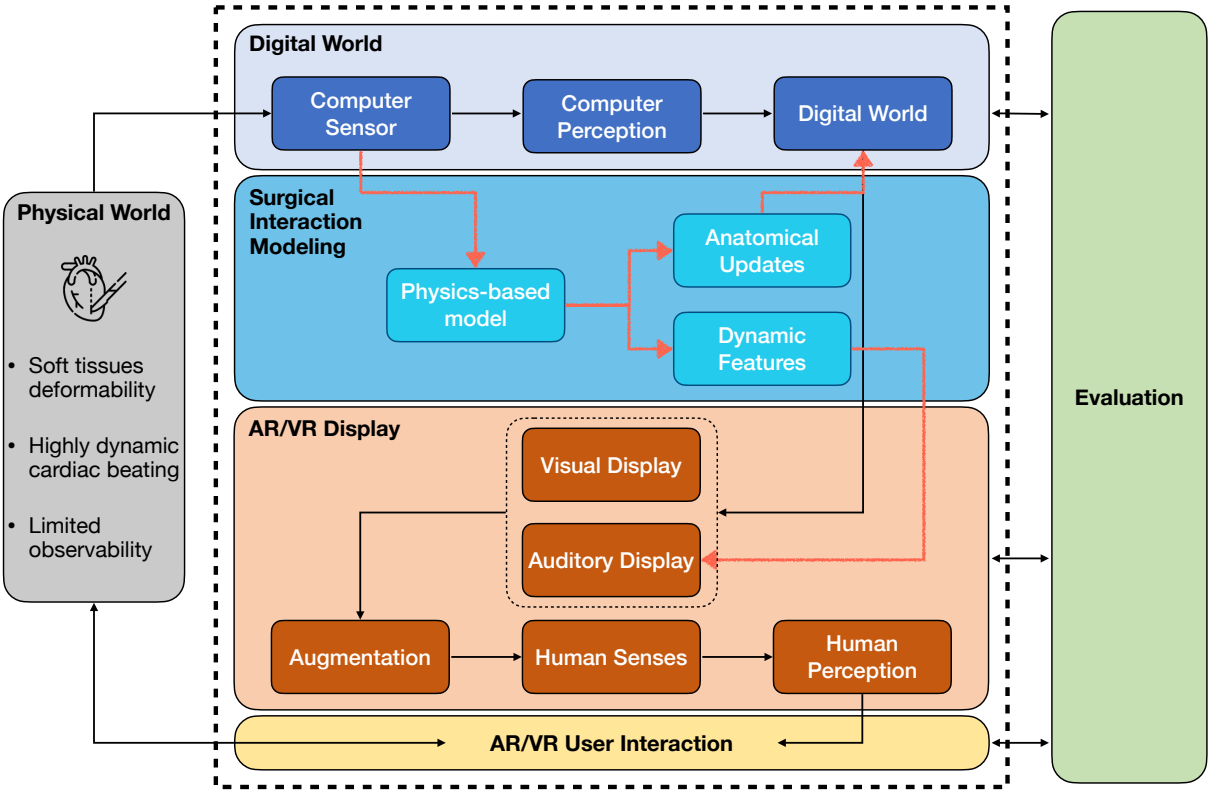


Figure 3.1: Medical XR Framework [157] enhanced with Surgical Interaction Modeling layer. The framework extension introduces a physics-based model of tool-tissue interaction positioned between the Physical World and the traditional Digital World representation. This model processes sensor-derived observations from the Physical World (catheter shape, instrument tracking) and generates two distinct outputs: (1) *Anatomical Updates* that refine the Digital World representation by predicting unobservable tissue states (vessel deformations, temporal cardiac motion), maintaining spatial correspondence under observational constraints; and (2) *Interaction Features* that encode dynamic properties of the surgical interaction (proximity to moving structures, contact forces, tissue resistance) and directly feed both Visual Display and Auditory Display modalities within the AR/VR Display layer. This dual-output architecture addresses both the spatial challenge (predicting what cannot be directly observed) and the perceptual challenge (delivering high-temporal-resolution guidance through complementary sensory channels). The Physical World characteristics particular to cardiovascular PIs, soft tissue deformability, highly dynamic cardiac beating, and limited observability, motivate the need for this physics-based interaction modeling approach.

physics of the interaction being navigated. This addresses the challenge of how the digital representation is presented to the user through adequate display modalities.

Consequently, physics-based modeling of surgical interactions, not merely anatomical modeling or updating, offers a unifying framework that addresses both spatial prediction under observational constraints (through anatomical updates to the Digital World) and temporal information delivery under perceptual constraints (through interaction features feeding display modalities), directly aligning with the Aim of this Thesis (Section 1.1).

This thesis addresses surgical interaction modeling through three distinct contributions in the field of cardiovascular percutaneous interventions. Each contribution demonstrates a different aspect of physics-based interaction modeling within the Medical XR Framework, addressing specific clinical use cases.

4 |

Enhanced Catheter Monitoring in Transcatheter Procedures

This chapter is based on: Ruozzi, V., Regazzo, G. B., Palumbo, M. C., Beckers, W.-A., Ourak, M., Zhang, X., Perico, F., Caimi, A., Vander Poorten, E., & Votta, E. (2026). Catheter monitoring in intelligent endovascular navigation systems: Interactive simulations and mixed reality for enhanced navigational awareness – International Journal of Computer Assisted Radiology and Surgery, 2026, 1-10. This work was conducted in collaboration with the Robot Assisted Surgery Group (RAS), KU Leuven.

Transcatheter procedures exemplify the challenges inherent to cardiovascular PIs introduced in Chapter 1. Access is typically achieved through percutaneous vascular puncture, most commonly via the femoral artery or vein, from which catheters navigate the vascular bed to deliver therapeutic devices, including valve prostheses, stents, and embolization coils, to the target anatomical site. These procedures have become established alternatives to open surgery [128], yet they remain technically demanding and require substantial operator expertise [5].

The central challenge lies in endovascular navigation: the controlled advancement and steering of flexible instruments through patient-specific vascular anatomy. Catheters and guidewires must traverse long and tortuous pathways while maintaining safe clearance from vessel walls. However, the diameter mismatch between large delivery systems and the vessels they navigate creates inherent collision risk. Commercially available sheaths can reach diameters of 8.0 mm for procedures such as MitraClip, approaching or exceeding the diameter of the femoral and iliac veins through which they navigate. Prolonged frictional contact or excessive mechanical force can lead to endothelial trauma, dissection, or perforation, with severe clinical consequences including stroke, arrhythmia, and vascular rupture [126, 131].

As discussed in Section 2.1.1, fluoroscopy provides excellent visualization of radiopaque instruments but does not depict vessel walls or their deformation during catheter manipulation. Consequently, catheter-vessel interactions and the mechanical forces that govern tissue damage remain invisible to the operator. While modern robotic platforms and force-sensing catheters attempt to address these limitations, real-time monitoring of contact forces and vessel deformation remains insufficient.

This observability gap motivates the work presented in this chapter.

4.1. Biomechanical Modeling of Catheter-Vessel Interaction

Traditional biomechanical approaches to modeling catheter navigation rely on FEM to predict vascular deformations induced by instrument insertion. These methods provide physically grounded predictions of vessel wall displacement and tool-tissue interaction forces by solving the governing equations of continuum mechanics under appropriate boundary and initial conditions.

The mechanical problem itself presents inherent numerical challenges: guidewire kinking, geometric instabilities arising from large deformations and rotations, and contact singular-

ities at sharp anatomical curvatures can cause convergence difficulties or solution failure in adverse anatomies [64].

Simulation frameworks for endovascular procedures typically reconstruct patient-specific vascular geometry from preoperative CT imaging, followed by volumetric mesh generation for both the vascular structure and the catheter or guidewire. Both instrument and vessel are discretized into finite elements, enabling the mechanical coupling between the two deformable bodies to be captured through contact formulations. The mechanical models incorporate nonlinear hyperelastic constitutive laws for the arterial wall, prestressing effects induced by blood pressure, and mechanical support from surrounding organs and skeletal structures [64]. Explicit time integration schemes enable robust handling of the complex contact dynamics, geometric nonlinearities, and large deformations that characterize instrument-vessel interaction. Validation studies against intraoperative imaging data have demonstrated encouraging predictive capability [65].

However, these high-fidelity simulations impose substantial computational demands. The number of elements used to discretize the catheter and vessel directly impacts both simulation accuracy and computational cost: finer meshes improve geometric resolution and contact detection but substantially increase the Degrees of Freedom (DOF) to be solved at each time step. Explicit time integration requires small time steps to maintain numerical stability, while contact detection and resolution between the two deformable meshes further increase computational cost. A single simulation can require several hours of computation time on workstation hardware, with the total cost scaling with mesh refinement and contact complexity [64].

Beyond computational cost, prediction fidelity critically depends on accurate specification of BCs that represent external mechanical constraints. Mechanical boundary conditions representing perivascular tissue, organ support, and dynamic cardiac motion are not directly observable from standard clinical imaging and must be estimated through calibration. Iterative calibration procedures that minimize the distance between boundaries extracted from dynamic cine-MRI sequences and simulation-derived vessel boundaries can substantially improve prediction accuracy, but require time-resolved imaging data and add further computational overhead [62]. This calibration challenge underscores a fundamental limitation: the unobservable parameters governing tissue support and anatomical motion significantly influence model predictions yet remain difficult to estimate reliably from standard imaging alone.

These computational characteristics fundamentally limit the applicability of traditional FEM approaches to intraoperative guidance scenarios. As established in Section 2.3.2,

surgical navigation systems require update rates compatible with human perception and interaction, typically demanding simulation updates at least around 25 Frames per Second (FPS) for visual feedback. The multi-hour computation times of high-fidelity models are incompatible with these temporal constraints.

Achieving interactive biomechanical simulation therefore requires substantial adaptation of both modeling strategies and numerical methods. As discussed in Section 2.3.2, optimized FEM solvers and constraint-based formulations enable real-time surgical interaction by rethinking numerical methods and the conceptual framework for tissue mechanics and device-tissue contact [50]. Within this paradigm, reduced-complexity models, such as beam theory for catheter representation combined with linear elastic vessel formulations, have enabled interactive simulations suitable for real-time interaction in training contexts [166], albeit with reduced fidelity compared to the full nonlinear formulations and refined meshes employed in offline prediction.

This trade-off between model complexity and computational performance remains central to the development of interactive surgical simulation systems, as will be addressed in the strategy presented in this chapter.

4.2. Contact Modeling Approaches

Contact modeling strategies for real-time FEM-based surgical simulations offer different trade-offs between computational speed, numerical stability, and physical accuracy [204]. The two dominant approaches are *penalty methods* and *Lagrange multiplier methods*.

Penalty methods approximate contact constraints by introducing artificial stiffness forces proportional to the penetration depth between contacting bodies. When two surfaces interpenetrate by a distance δ, a repulsive force $F = k_p\delta$ is applied, where k_p is a penalty stiffness parameter. This approach requires no modification to the system matrix structure and adds minimal computational overhead. The penalty force is simply added to the right-hand side of the equation system. However, penalty methods suffer from a fundamental limitation: selecting an appropriate k_p value involves a critical trade-off between accuracy and stability. Large penalty stiffness values reduce penetration error but worsen the condition number of the system matrix, often causing convergence problems and numerical instabilities that require prohibitively small timesteps for stable integration. Small penalty stiffness values maintain computational efficiency but permit unrealistic interpenetration. The selection of k_p is problem-dependent and heavily depends on the ratio of material stiffness between the contacting objects [204], making parameter tuning application-specific and non-transferable across scenarios.

Lagrange multiplier methods enforce contact constraints exactly by treating contact forces as unknown variables solved simultaneously with the system. The equation system is augmented to include constraint contributions, with contact forces represented by Lagrange Multiplier (LM) λ that impose the impenetrability condition. This formulation eliminates parameter tuning entirely and guarantees zero penetration at convergence, providing physically accurate contact behavior regardless of material properties or geometric complexity. The computational cost, however, is substantially higher: in addition to the body systems, an additional dense linear system of dimension equal to the number of active constraints must be assembled and solved at each timestep, and each contact requires further back-substitutions on the previously factorized body matrices.

Studies on probe-breast interaction have demonstrated that LM methods achieve superior stability compared to penalty approaches, particularly at large deformations where nonlinear effects become significant. However, computation time can increase by an order of magnitude when using direct solvers for the augmented system [204].

The catheter-vessel interaction problem presents precisely the scenario where LM formulations become necessary despite their computational overhead. Catheters are slender, highly flexible structures prone to buckling and kinking under compressive loads. Vessels exhibit complex tortuosity, creating intricate mutual loading conditions between the catheter and vessel walls. The coupling of these geometries represents a mechanically unstable configuration: small changes in contact location or force magnitude can trigger buckling modes, sudden geometric reconfigurations, or catheter instabilities. Penalty-based contact, with its inherent parameter sensitivity, cannot reliably stabilize such interactions. For our application, allowing even small interpenetration could lead to physically implausible vessel deformation due to catheter contact, while attempting to prevent penetration through high penalty stiffness would trigger the numerical instabilities documented in the literature [204].

Consequently, the framework presented in this chapter adopts a LM formulation for contact modeling, accepting the increased computational cost as necessary to achieve numerical stability and physical accuracy when complex deformable geometries interact. The computational implications of this choice are analyzed quantitatively in Section 4.2.2.

4.2.1. Lagrange Multiplier Formulation

This section describes the LM formulation implemented for modeling catheter-vessel contact when an implicit time-integration scheme is adopted to solve the dynamic equilibrium

equations at each timestep. Without contact constraints, the problem is formulated as:

$$\mathbf{A}\,\boldsymbol{\Delta v} = \mathbf{b}, \tag{4.1}$$

where $\mathbf{A}$ is the system matrix (combining mass, damping, and stiffness contributions), $\boldsymbol{\Delta v}$ represents incremental velocities, and $\mathbf{b}$ gathers external and inertial forces.

Contact constraints are handled through LMs. For a pair of objects in contact, a scalar LM λ represents the magnitude of the contact force. The constraint problem is decomposed into two sequential steps following the SOFA framework [1]:

Free Motion Each object is solved independently with $\boldsymbol{\lambda} = 0$, computing the unconstrained velocity change:

$$\mathbf{A}\,\boldsymbol{\Delta v}^{free} = \mathbf{b}. \tag{4.2}$$

Constraint Correction Constraints are enforced with $\mathbf{b} = 0$, computing the corrective velocity change:

$$\mathbf{A}\,\boldsymbol{\Delta v}^{corr} = dt\,\mathbf{H}^T\boldsymbol{\lambda}, \tag{4.3}$$

where $\mathbf{H}$ is the constraint Jacobian, mapping configuration-space velocities into constraint-space velocity violations, so that its transpose $\mathbf{H}^T$ maps the constraint forces (the multipliers $\boldsymbol{\lambda}$) back into generalized forces in configuration space, and dt is the timestep.

For a contact between catheter point P and vessel point Q, the global constraint Jacobian $\mathbf{H} = [\mathbf{H}_P, \mathbf{H}_Q]$ is partitioned into the two body blocks, defined with appropriate signs so that the relative motion across the contact is correctly recovered. The constraint violation in velocity then reads:

$$\dot{\delta} = \mathbf{H}_P\mathbf{v}_P^{free} + \mathbf{H}_Q\mathbf{v}_Q^{free}, \tag{4.4}$$

where $\mathbf{v}_P^{free}$ and $\mathbf{v}_Q^{free}$ are the free-motion velocities. The penetration depth δ returned by collision detection provides the initial position-level violation.

The displacement–force relationship is expressed through the compliance matrix. Substituting the corrective motion ($\boldsymbol{\Delta v}_i^{corr} = dt\,\mathbf{A}_i^{-1}\mathbf{H}_i^T\boldsymbol{\lambda}$) into the constraint equation yields:

$$\dot{\delta} = \mathbf{H}_P\mathbf{v}_P^{free} + \mathbf{H}_Q\mathbf{v}_Q^{free} + dt\left[\mathbf{H}_P\mathbf{A}_P^{-1}\mathbf{H}_P^T + \mathbf{H}_Q\mathbf{A}_Q^{-1}\mathbf{H}_Q^T\right]\boldsymbol{\lambda}. \tag{4.5}$$

[1] See the SOFA documentation

Defining the compliance matrix:

$$\mathbf{W} = dt \left[\mathbf{H}_P \mathbf{A}_P^{-1} \mathbf{H}_P^T + \mathbf{H}_Q \mathbf{A}_Q^{-1} \mathbf{H}_Q^T\right], \tag{4.6}$$

and collecting all m contact pairs, the constraint system becomes:

$$\mathbf{W}\boldsymbol{\lambda} + \dot{\delta}^{free} = 0, \tag{4.7}$$

where $\mathbf{W} \in \mathbb{R}^{m \times m}$ is the dense compliance matrix in constraint space and $\dot{\delta}^{free}$ is the constraint-velocity violation vector arising from the free motion. For unilateral contacts, Eq. 4.7 is supplemented by the Signorini complementarity conditions $\dot{\delta} \geq 0$, $\boldsymbol{\lambda} \geq 0$, $\boldsymbol{\lambda}^T \dot{\delta} = 0$.

This system is solved iteratively via a projected Gauss–Seidel algorithm, which enforces the complementarity conditions imposed by Signorini's law for unilateral contacts. Once $\boldsymbol{\lambda}$ is determined, the corrected velocities are:

$$\mathbf{v}_P = \mathbf{v}_P^{free} + dt\, \mathbf{A}_P^{-1} \mathbf{H}_P^T \boldsymbol{\lambda}, \quad \mathbf{v}_Q = \mathbf{v}_Q^{free} + dt\, \mathbf{A}_Q^{-1} \mathbf{H}_Q^T \boldsymbol{\lambda}, \tag{4.8}$$

and positions are updated by integration:

$$\mathbf{x} = \mathbf{x}^{free} + dt\, \boldsymbol{\Delta v}^{corr}. \tag{4.9}$$

The LM formulation introduces significant computational overhead, as it requires building compliance matrices and solving an additional linear system per timestep.

4.2.2. Computational Cost Analysis

The computational cost of the LM-based contact resolution depends critically on the dimensionality of the interacting models and the number of active contact constraints. This section analyzes the cost structure for a general catheter-vessel interaction scenario.

System Dimensionality

Let n_c denote the number of catheter DOFs, n_v the number of vessel DOFs, and m the number of active contact pairs detected per timestep. For the catheter-vessel interaction:

Catheter Model The catheter is modeled using the SOFA plugin *BeamAdapter*, which implements a corotational Timoshenko beam formulation [222]. In this framework, each

beam element connects two nodes, each carrying 6 DOFs: three translational components and three rotational components. The elementary stiffness matrix $\mathbf{K}_e$ is a 12×12 symmetric matrix relating the nodal velocities and angular velocities at both ends of the element to the corresponding forces and torques:

$$\delta\mathbf{F}(\mathbf{q}) = \left[\mathbf{\Lambda}^T(\mathbf{q})\,\mathbf{K}_e\,\mathbf{\Lambda}(\mathbf{q})\right]\delta\mathbf{v}, \tag{4.10}$$

where $\mathbf{\Lambda}(\mathbf{q})$ is the direction-cosine transformation matrix mapping local element coordinates to the global frame, and $\delta\mathbf{v}$ represents nodal velocity increments. The corotational assumption confines material deformations to a local element frame, allowing the model to handle geometrically nonlinear behavior arising from large deflections of the catheter during navigation.

The geometry of the beam is supported by cubic Bézier splines, interpolated segment by segment between adjacent frames via *WireBeamInterpolation*. Forces and masses are computed by *AdaptiveBeamForceFieldAndMass*, while *AdaptiveBeamMapping* transfers forces and contact constraints from the collision surface back to the beam DOF.

Vessel Model The vessel is modeled using a first-order tetrahedral finite element formulation. Each node carries 3 translational velocity DOFs. Vessel mechanics are simulated using a corotational formulation. At each timestep, the system matrices (stiffness $\mathbf{K}$ and mass $\mathbf{M}$) are assembled and contribute to the implicit system in Eq. 4.1 for updating vessel velocity. The SOFA framework implements this via *TetrahedralCorotationalFEMForceField* and *MeshMatrixMass* components. Vessel configuration, as set of nodal positions, is then obtained through time integration of the computed velocities.

Cost Structure of Lagrange Multiplier Contact Resolution

The computational cost per timestep can be decomposed into four contributions, corresponding to the free motion, the assembly of the compliance matrix, the iterative resolution of the constraint system, and the application of the corrective motion:

1. Free Motion Phase Each object is solved independently with $\boldsymbol{\lambda} = 0$ by solving:

$$\mathbf{A}_c\,\Delta\boldsymbol{v}_c^{free} = \mathbf{b}_c, \quad \mathbf{A}_v\,\Delta\boldsymbol{v}_v^{free} = \mathbf{b}_v. \tag{4.11}$$

With a sparse direct solver (e.g., LDL^T factorization), the cost of this step grows superlinearly with the number of DOFs of each body. We denote the combined free-motion

cost generically as:

$$\mathcal{C}_{free} = \mathcal{O}(n_c^{\alpha_c} + n_v^{\alpha_v}), \tag{4.12}$$

where $\alpha_c, \alpha_v \geq 1$ are problem-dependent exponents associated, respectively, with the catheter and the vessel system.

2. Compliance Matrix Construction For each active contact, the compliance matrix (Eq. 4.6) requires computing:

$$\mathbf{W} = dt \left[\mathbf{H}_c \mathbf{A}_c^{-1} \mathbf{H}_c^T + \mathbf{H}_v \mathbf{A}_v^{-1} \mathbf{H}_v^T \right]. \tag{4.13}$$

Each term $\mathbf{H}_i \mathbf{A}_i^{-1} \mathbf{H}_i^T$ is built by reusing the already-computed factorization of $\mathbf{A}_i$ to back-substitute one right-hand side per active constraint. Denoting by $\beta_c, \beta_v \geq 1$ the corresponding problem-dependent back-substitution exponents, the compliance assembly cost can be expressed as:

$$\mathcal{C}_{compliance} = \mathcal{O}\big(m\,(n_c^{\beta_c} + n_v^{\beta_v}) + m^2\big)\,, \tag{4.14}$$

where the first term accounts for the m back-substitutions on each body and the second for the m^2 inner products needed to assemble the dense $m \times m$ matrix $\mathbf{W}$.

3. Constraint System Solution The constraint system $\mathbf{W}\boldsymbol{\lambda} + \dot{\delta}^{free} = 0$ (Eq. 4.7) is solved via Gauss-Seidel iteration. The compliance matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$ is dense due to mechanical coupling through both bodies. Each Gauss-Seidel iteration costs $\mathcal{O}(m^2)$. With k iterations required for convergence:

$$\mathcal{C}_{constraint} = \mathcal{O}(k \cdot m^2). \tag{4.15}$$

4. Corrective Motion Application Once $\boldsymbol{\lambda}$ is determined, corrective velocities are computed:

$$\boldsymbol{\Delta v}_c^{corr} = dt\, \mathbf{A}_c^{-1} \mathbf{H}_c^T \boldsymbol{\lambda}, \quad \boldsymbol{\Delta v}_v^{corr} = dt\, \mathbf{A}_v^{-1} \mathbf{H}_v^T \boldsymbol{\lambda}. \tag{4.16}$$

In practice this step reuses the per-constraint vectors $\mathbf{A}_i^{-1} \mathbf{H}_{i,k}^T$ already computed during the assembly of $\mathbf{W}$, so that no further factorization is needed and the correction reduces to a sparse linear combination over the m active constraints:

$$\mathcal{C}_{correction} = \mathcal{O}(m \cdot (n_c + n_v)). \tag{4.17}$$

Total Cost per Timestep Combining the four contributions, the overall per-timestep computational cost can be summarised as:

$$\mathcal{C}_{total} = \mathcal{O}\big(n_c^{\alpha_c} + n_v^{\alpha_v} \;+\; m\,(n_c^{\beta_c} + n_v^{\beta_v}) \;+\; k\,m^2 \;+\; m\,(n_c + n_v)\big)\,, \tag{4.18}$$

with $\alpha_i \geq \beta_i \geq 1$ being problem-dependent exponents larger than one.

This analysis reveals that the cost scales super-linearly with both bodies' DOFs, highlighting the reward in computational cost if some of the DOFs are eliminated. This computational challenge motivates the framework design and implementation described in the following section.

4.3. Methods

This section describes the design, implementation and validation of the proposed framework through in-vitro testing.

The framework relies on three key **design choices** to enable interactive biomechanical monitoring: (1) sensor-driven simulation that eliminates the catheter from the mechanical system by reconstructing its configuration from real-time shape sensing, (2) continuum-embedded vessel representation that provides computationally efficient boundary conditions, and (3) side-by-side XR visualization that presents predicted deformations within the surgeon's field of view.

The **implementation** of the framework into the in-vitro set-up integrates four components: (i) a fiber-optic and electromagnetic sensorized catheter for real-time shape reconstruction, (ii) a finite element model of the vessel for interactive deformation simulation, (iii) hardware infrastructure for sensor data acquisition and coordinate frame alignment, and (iv) a XR interface for visualization. These components operate within a three-step workflow, *Initialization*, *Registration*, and *Runtime*, illustrated in Figure 4.1.

The complete experimental setup is described first, followed by a description of the three workflow steps. Subsequently, two sets of experiments are presented. Section 4.3.5 describes **computational benchmarking analyses** conducted to systematically select numerical parameters governing the simulation, including mesh discretization, mechanical properties, and integration timestep. Section 4.3.6 presents the **validation experiments** assessing the accuracy and performance of the simulated vessel deformation.

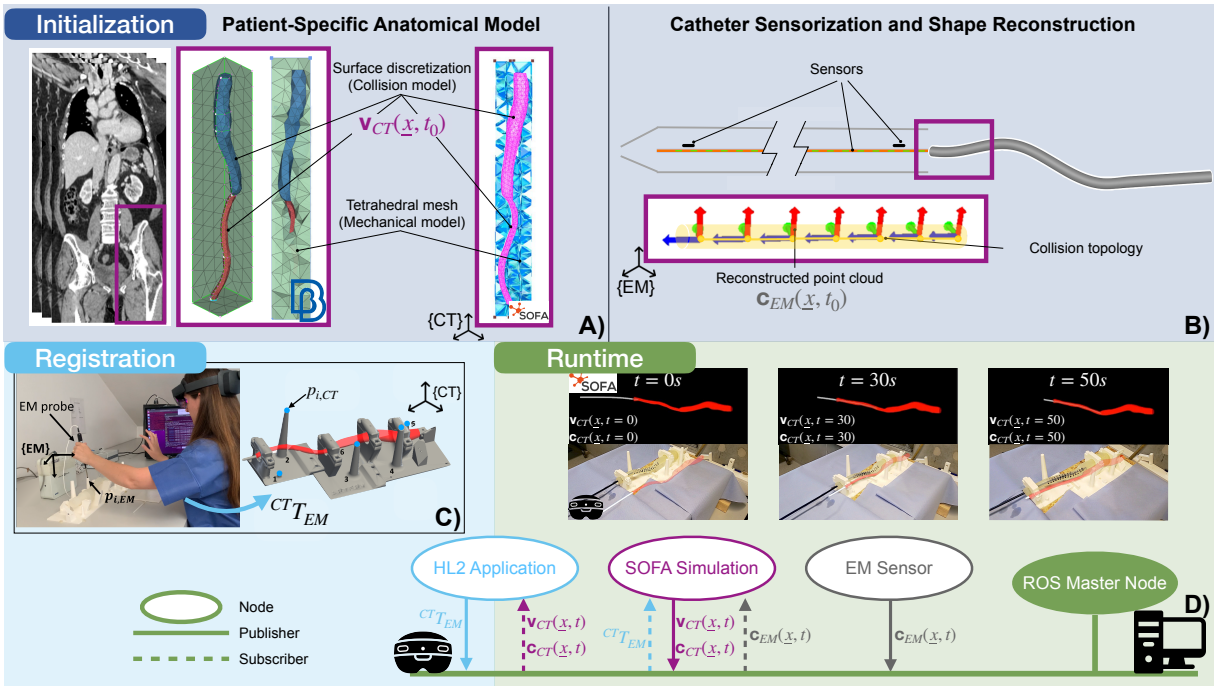


Figure 4.1: Schematic overview of the proposed framework. An *Initialization* step pre-processes vessel (A) and catheter (B) models and instantiates the interactive simulation. A *Registration* step aligns sensor ($\{EM\}$) and anatomical ($\{CT\}$) frames (C). During *Runtime*, the interactive simulation continuously updates the XR visualization (D).

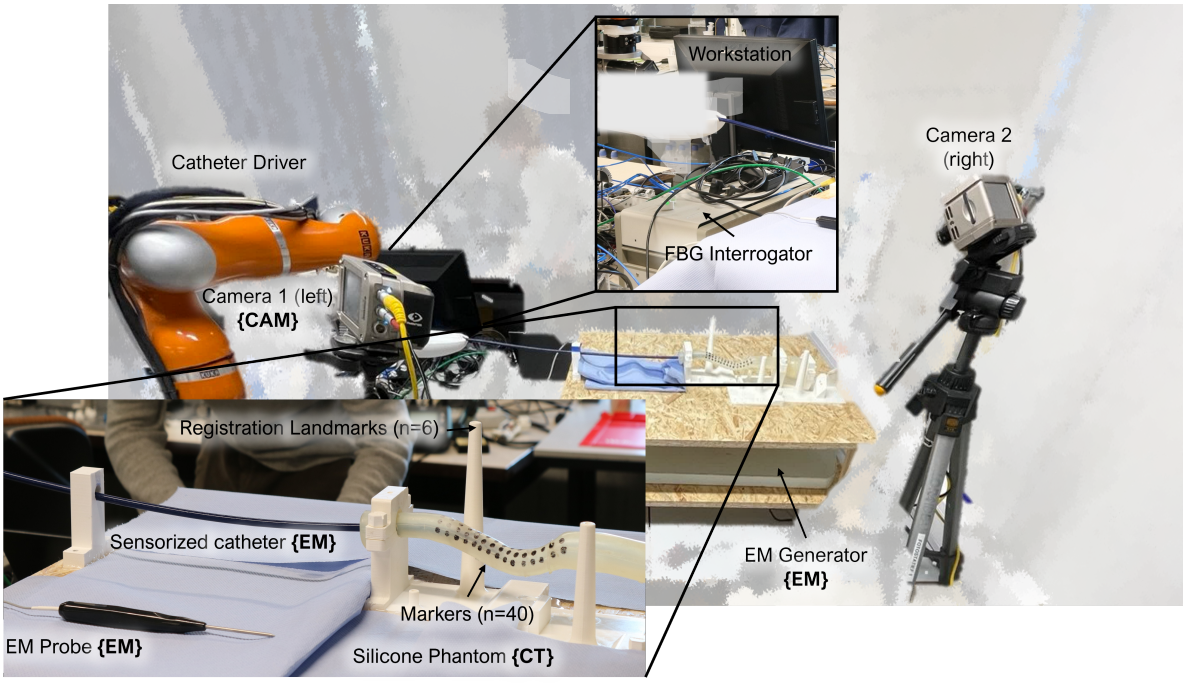


Figure 4.2: Experimental setup for implementing and testing the framework. A robotic driver inserted the sensorized catheter into a silicone phantom replicating the vessel anatomy. Two high-speed cameras tracked surface-marker displacements on the phantom wall, while the corresponding FEM nodes were identified by applying the transformation ${}^{CT}\mathbf{T}_{EM}$ to the point cloud acquired at rest using the EM probe.

4.3.1. In-Vitro Setup

The setup (Figure 4.2) used a silicone phantom of the venous lumen from the femoral vein to the inferior vena cava (45.6 cm), derived from contrast-enhanced CT scans. Veins were cast with 4.5 mm walls, mounted on a 3D-printed support with six markers, and marked with 40 surface points for deformation tracking. A 7.4 mm catheter diameter was instrumented with two 5 DOF EM sensors (NDI - Northern Digital Inc., ON, Canada) and a 2 m-long multicore FBG sensor (38 gratings, 10 mm spacing; FBGS, Geel, Belgium). EM sensors were placed at the tip and 18 cm proximally [69]. Shape was reconstructed in real-time (40 Hz) using twist-compensated calibration and the helical extension method [6] for 3D reconstruction, fused with EM data for absolute positioning and resampled at 78 points ($\mathbf{c}_{EM}(\underline{x}, t) = \{\mathbf{c}_{iEM}(t)\}_{i=1}^{78}$) over 37 cm. Additional components included a robotic driver [7], FBG interrogator, EM generator, EM Aurora® NDI Probe (Aurora, Northern Digital Inc., Waterloo, ON, Canada), Hololens 2™ (Hololens 2™ (H2)) (Microsoft Inc.) HMD, and two high-speed stereo cameras (Phantom Miro 3, Vision Research Inc., Wayne, NJ, USA). Cameras tracked marker displacement, while the Aurora probe provided reference points for registration.

4.3.2. Initialization

Virtual avatars of the CT-derived vein (Figure 4.1A) and the catheter (Figure 4.1B) were generated to initialize the FEM.

Vein The vessel anatomy, reconstructed from CT imaging as the inner lumen surface without vessel wall thickness, was embedded within a deformable continuum to represent the mechanical constraints imposed by surrounding tissues. The segmented lumen was imported into ANSA (Beta CAE Systems, USA) and enclosed within a hexahedral *BBox* expanded by 10% along the main vein axis (Z) and 20% along the transverse directions (X and Y) relative to the anatomical reference frame $\{CT\}$. The *BBox* employs a homogenization strategy wherein the continuum material represents the composite mechanical behavior of vessel wall and adjacent soft tissues through linear elastic properties. Surface discretization was performed using triangular elements with characteristic lengths of 8 mm (femoral-iliac segment), 15 mm (inferior vena cava), and 30 mm (*BBox* outer boundary). The bulk continuum was discretized into linear tetrahedral elements with maximum aspect ratio of 5. Material properties were assigned as: density $\rho = 1000$ kg/m^3, Young's modulus $E = 1$ MPa, and Poisson's ratio $\nu = 0.36$. These material parameters were selected based on soft tissue mechanical properties as detailed in Section 4.3.5. A corotational formulation accounts for geometric nonlinearity arising from large deformations

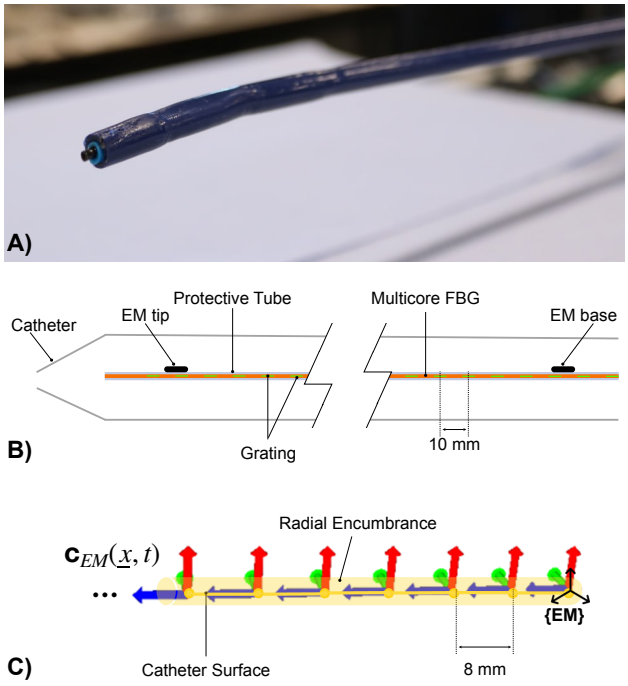


Figure 4.3: A 7.4 mm catheter **(A)** housed a multi-core FBG fiber and two EM sensors within its central lumen **(B)**. The sensor-based reconstructed catheter centerline, expressed in the $\{EM\}$ reference frame as $\mathbf{c}_{\mathrm{EM}}(\underline{x}, t)$, was uniformly discretized into 78 points. Each point served as the center of a rigid sphere, and consecutive points were connected by cylinders of 7.4 mm diameter, producing a piecewise-continuous geometric avatar of the catheter outer surface for collision and contact handling **(C)**.

induced by catheter contact. At each timestep, the implicit system (Eq. 4.1) is solved to update vessel configuration via the LM formulation (Section 4.2.1). Mesh resolution and discretization timestep (dt) were selected through the sensitivity analyses presented in Section 4.3.5.

Catheter The discretized centerline of the catheter avatar consisted in the set of 78 points $\mathbf{c}_{\text{EM}}(\underline{x}, t) = \{\mathbf{c}_i(t)\}_{i=1}^{78}$ whose coordinates were in the electromagnetic frame (i.e., $\{EM\}$) and continuously monitored via EM and FBG sensors. Consistently with the dimensions of the real catheter, a rigid sphere with 7.4 mm diameter was centered at each point $\mathbf{c}_i(t)$, and a cylinder with 7.4 mm diameter was aligned with the segment running between consecutive points $\mathbf{c}_i(t)$ and $\mathbf{c}_{i+1}(t)$, with $i = 1, ..., 77$ (Figure 4.3).

4.3.3. Registration

The rigid transformation matrix ${}^{CT}\mathbf{T}_{EM}$ aligning the $\{EM\}$ to the $\{CT\}$ reference frame was computed via a marker-based Singular Value Decomposition (SVD) registration (Section 2.1.3). Six pillars were integrated into the phantom support (Figure 4.2) around the anatomical region of interest to span the EM field generator's working volume across different heights and served as registration markers. Their spatial coordinates in $\{CT\}$ ($\mathbf{p}_{i,CT}$, $i = 1, ..., 6$) were known a priori, while corresponding points in $\{EM\}$ ($\mathbf{p}_{i,EM}$, $i = 1, ..., 6$) were acquired by placing the EM probe into dedicated holders at the top of each pillar (Figure 4.1C). The resulting registration yielded a Root Mean Square Error (RMSE) of 0.982 mm. Registration was managed through a XR interface developed in Unity3D (Unity Technologies, version 2020.3.25LTS) with the Mixed Reality Toolkit (MRTK v2.7.3), deployed to HL2.

4.3.4. Runtime

After initialization and registration, catheter point coordinates in the $\{CT\}$ frame ($\mathbf{c}_{CT}(\underline{x}, t)$) were streamed to the SOFA framework [54], with

$$\mathbf{c}_{CT}(\underline{x}, t) = {}^{CT}\mathbf{T}_{EM}\, \mathbf{c}_{EM}(\underline{x}, t).$$

Vessel-catheter interaction was modeled using a corotational FEM formulation with a Lagrange multiplier (LM)-based collision model applied to both lumen and catheter surfaces. Two constraint sections were defined at the proximal and distal $BBox$ faces. For visualization, a 2 mm triangular mesh was mapped to the lumen surface, and its vertex coordinates ($\mathbf{v}_{CT}(\underline{x}, t)$) were streamed at each dt to the XR application for updating.

Data transfer from sensors to simulation and from simulation to the XR application was handled via Robotic Operating System (ROS) inter-node communication (Figure 4.1D).

4.3.5. Computational Benchmarking

Before running the validation experiments, numerical parameters governing the interactive simulation, mesh resolution, material properties, and integration timestep, were systematically investigated using a controlled mock insertion scenario. This approach ensured consistent test conditions across all parameter variations, isolating the influence of each numerical choice on simulation performance and accuracy.

A fictitious catheter trajectory was generated using Non-Uniform Rational B-Spline (NURBS) interpolating the vessel centerline, with simulated insertion at 10 mm/s (Figure 4.4). Although higher than the 5 mm/s speed adopted in the validation experiments described in the subsequent section, this velocity was chosen to stress-test the simulation under challenging conditions characterized by significant inertial effects. The trajectory incorporated variations from the centerline in three-dimensional vessel curvature to evaluate the stability of the LM contact model under geometrically demanding configurations.

Figure 4.4A demonstrates that the LM formulation maintains stability throughout the insertion, correctly computing the deformed vessel configuration despite the complex trajectory. Figure 4.4B shows the displacement field magnitude for the baseline configuration ($E = 1.0$ MPa, $\nu = 0.36$), which serves as the reference for all sensitivity analyses described below. A region of interest (Region of Interest (ROI)) was defined for localized displacement analysis, while global deformation was assessed over the full vessel lumen.

Three investigations were conducted using this test scenario: (i) mesh sensitivity analysis to determine the optimal spatial discretization balancing accuracy and computational efficiency, (ii) mechanical parameter sensitivity to assess the influence of Young's modulus and Poisson's ratio on displacement predictions, and (iii) time integration scheme validation to verify that adaptive timestep control maintains accuracy. For all analyses, vessel configurations were compared when the catheter tip reached the distal end of the iliac bifurcation (configuration $\Gamma_{\bar{t}}$, where $\bar{t}$ denotes the time at which the catheter tip reached this specific anatomical location). Simulations were executed on a workstation equipped with an Intel Core i7-10750H processor and an NVIDIA GeForce GTX 1660Ti GPU.

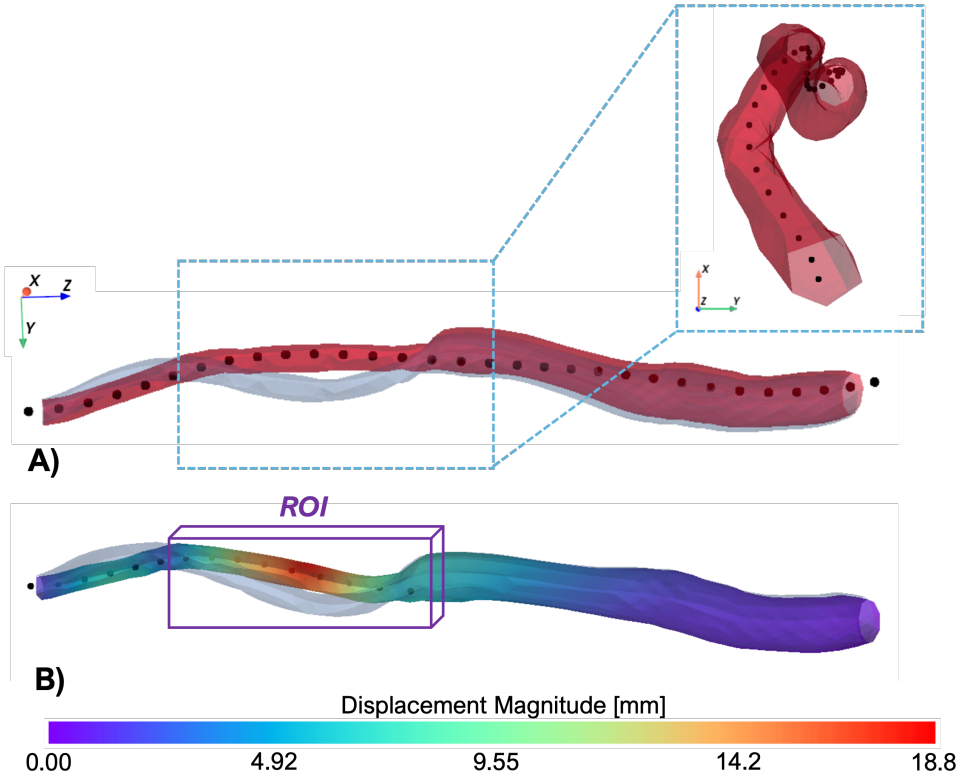


Figure 4.4: Mock insertion trajectory for computational benchmarking. A fictitious catheter trajectory generated via NURBS interpolation of the vessel centerline was used to simulate insertion at 10 mm/s. **(A)** The LM contact model maintains stability and correctly computes the deformed vessel configuration induced by the complex trajectory. **(B)** Displacement field magnitude for the baseline configuration ($E = 1.0$ MPa, $\nu = 0.36$), serving as the reference for all sensitivity analyses. The region of interest (ROI) indicates the zone used for localized displacement metrics, while global metrics were computed over the full vessel lumen.

Mesh Sensitivity Analysis

The refinement of the *BBox* mesh impacts both computational cost and simulation accuracy. A sensitivity analysis was conducted using four distinct mesh resolutions to identify the optimal compromise between computational efficiency and displacement accuracy on the vessel lumen.

Displacement variation was quantified by the 95$^{\text{th}}$ percentile of the nodal displacement magnitude on the vessel lumen ($|\mathbf{U}_{95}|$) when the catheter tip reached the distal end of the iliac bifurcation (configuration $\Gamma_{\bar{t}}$). Computational cost was assessed through total accumulated machine time (T_{CPU}) and average rendered frames per second (FPS). Table 4.1 summarizes the characteristics and performance of the four tested meshes.

Table 4.1: Mesh sensitivity analysis results. T_{CPU} is total accumulated machine time, and $|\mathbf{U}_{95}|$ is the 95$^{\text{th}}$ percentile of displacement magnitude at configuration $\Gamma_{\bar{t}}$.

Mesh	**Elements**	**Nodes**	T_{CPU} **(s)**	**FPS (Hz)**	$\lvert\mathbf{U}_{95}\rvert$ **(mm)**
Very Coarse	1106	289	216.0	13.3	3.05
Coarse	2003	546	351.8	9.7	3.02
Medium	3659	946	651.3	6.2	3.02
Fine	8021	1969	1557.8	3.1	2.99

Results demonstrated that computational cost scaled steeply with mesh resolution: T_{CPU} increased 7-fold from *Very Coarse* to *Fine*, while FPS dropped from 13.3 to 3.1 Hz. Conversely, displacement solutions exhibited negligible variation across all tested meshes, with $|\mathbf{U}_{95}|$ ranging between 2.99 and 3.05 mm (less than 2% variation). This convergence behavior arises from the global kinematic constraints imposed by the boundary conditions, which dominate the deformation pattern and reduce sensitivity to local mesh refinement.

The *Coarse* mesh (2003 elements, 546 nodes) was selected for all subsequent simulations, providing an acceptable balance between computational efficiency (9.7 Hz) and displacement accuracy (within 1% of the finest mesh solution). While below the ideal interactive threshold of 15 Hz or higher, this performance is compatible with the adaptive timestep strategy described below.

Mechanical Parameters

The mechanical parameters Young's modulus $E = 1$ MPa and Poisson's ratio $\nu = 0.36$ specified in the Initialization (Section 4.3.2) constitute a baseline configuration selected as a compromise between the modulus of soft tissues (adipose and connective tissue) and that of the vessel wall within the deformation range of approximately 30%. Literature values for

Young's modulus of adipose tissue range from 0.12 to 0.25 MPa, and for connective tissue from 0.49 to 0.99 MPa [100]. Furthermore, similar deformable constraint characterizations report values on the order of 1 MPa [59].

The impact of these constitutive parameters on the computed vessel lumen displacement was investigated through sensitivity analysis. Both parameters were systematically varied within physiologically plausible ranges. Two metrics were defined to distinguish local response, which is more susceptible to contact modeling, from global response. The impact on local displacements was determined by monitoring only the nodes within the defined ROI (Figure 4.4) through the metric $\Delta\mathbf{U}$, defined as the absolute difference between the 95^{th} percentile of the displacement field norm of the i^{th} simulation with respect to the baseline:

$$\Delta\mathbf{U}_i = |\|\mathbf{U}\|_{95,i} - \|\mathbf{U}\|_{95,\text{baseline}}|$$

with $i = 1, \ldots, 23$. The global response was quantified by monitoring the entire vessel lumen through the Chamfer Distance (*ChDist*). Considering the node set at the baseline configuration Γ_{baseline} (denoting the reference configuration with baseline material parameters) as $\Gamma_{\text{baseline}} = \{x_p \in \mathbb{R}^3\}_{p=1}^n$ and that of each i^{th} simulation $\Gamma_i = \{x_q \in \mathbb{R}^3\}_{q=1}^n$, this metric is defined as the average distance between pairs of nearest neighbors:

$$\mathit{ChDist}(\Gamma_{\text{baseline}}, \Gamma_i) = \frac{1}{2n}\sum_{p=1}^{n} |x_p - \mathbf{NN}(x_p, \Gamma_{baseline})| + \frac{1}{2n}\sum_{q=1}^{n} |x_q - \mathbf{NN}(x_q, \Gamma_i)|$$

where $\mathbf{NN}(x, \Gamma) = \arg\min_{x' \in \Gamma} \|x - x'\|$ is the nearest neighbor function.

Results indicated moderate sensitivity to E, with local displacement deviations ranging from 0.06 to 0.56 mm and global Chamfer distances from 0.02 to 0.14 mm as E varied from 0.5 to 2.0 MPa. In contrast, variations in ν from 0.30 to 0.45 produced negligible changes, with $\Delta\mathbf{U} < 0.05$ mm and $\mathit{ChDist} < 0.01$ mm. These findings confirm that E should be selected based on tissue mechanical properties when available, while ν can be set to a standard near-incompressible value (0.36) without significantly affecting displacement predictions. Detailed sensitivity data are provided in Appendix A.

Time Integration Scheme

Once the spatial discretization and constitutive mechanical parameters had been defined, the numerical simulation configuration was largely complete. A final analysis was conducted to verify that the SOFA *Real-Time Controller*, which adaptively adjusts the integration timestep *dt* to maintain interactivity, does not compromise simulation accuracy.

The *Real-Time Controller* prevents accumulated machine time T_{CPU} from exceeding acceptable limits by dynamically increasing dt when individual iterations are computationally expensive. While this strategy maintains interactive performance, excessively large timesteps could degrade accuracy. Two simulations were therefore compared: one with a fixed timestep ($dt = 0.01$ s), and one using the *Real-Time Controller* with adaptive dt. Both simulations were evaluated at configuration $\Gamma_{\bar{t}}$ using the 95$^{\text{th}}$ percentile of displacement magnitude ($|\mathbf{U}_{95}|$) as the accuracy metric.

A difference of 0.218 mm was observed between the $|\mathbf{U}_{95}|$ values obtained with fixed and adaptive timestep schemes, corresponding to approximately 7% relative difference. This negligible deviation confirms that the *Real-Time Controller* can be safely employed to maintain interactive performance without sacrificing displacement accuracy.

4.3.6. In-vitro Experiments

The experiment was repeated three times. After registration, the EM Aurora® probe recorded the initial positions of all outer-surface markers, generating the reference point cloud $\mathbf{p}_{EM}(\underline{x}_i, t_0)$ with $i = 0, \ldots, 39$. Three points per clamp on the rigid 3D-printed base were also recorded with the probe to update the FEM constraint sections. The stereo cameras were calibrated using a checkerboard pattern. The robotic driver then advanced the catheter into the silicone phantom at 5 mm/s without active steering, a speed selected based on prior work [7] and commercial robotic catheter driver capabilities [213].

The start of catheter insertion triggered the synchronized acquisition of two datasets:

1. *experimental dataset*, stereo image sequences from the left and right cameras (Phantom Camera Control software, Vision Research Inc., USA);

2. *computational dataset*, i) a *rosbag* logging the 78 catheter points at 40 Hz vs. physical time; ii) an ASCII file containing the nodal coordinates of the catheter and vessel avatars as a function of the central processing unit (CPU) computation time (accumulated machine time (T_{CPU}), where T_{CPU} denotes the time required by the CPU to compute and update the current mesh state.

At the end of each experiment, 13 time-points spanning catheter motion were selected, with 8 s between consecutive points. For each time-point, the following quantities were measured:

1. interactive performance – comparing T_{CPU} from ASCII files with physical time (T_{PHYS}) from *rosbags*, using the one-to-one correspondence between real and simulated catheter points; HL2 rendering frames per second (FPS) was also monitored.

2. the accuracy of the simulated vessel deformation – markers on the outer surface of the silicone phantom were manually tracked in stereo images and reconstructed in 3D via triangulation [91], yielding $\mathbf{p}_{CAM}(\underline{x}, t)$. At $t = t_0$, a principal component analysis-based registration computed ${}^{EM}\mathbf{T}_{CAM}$ aligning $\mathbf{p}_{CAM}(\underline{x}, t_0)$ to $\mathbf{p}_{EM}(\underline{x}, t_0)$. This transformation was then composed with the transformation matrix ${}^{CT}\mathbf{T}_{EM}$ to express $\mathbf{p}_{CAM}(\underline{x}, t)$ in the $\{CT\}$ reference frame:

$$\mathbf{p}_{EXP}(\underline{x}, t) = {}^{CT}\mathbf{T}_{EM} \cdot {}^{EM}\mathbf{T}_{CAM} \cdot \mathbf{p}_{CAM}(\underline{x}, t)$$

Marker ground-truth displacements were computed from $\mathbf{p}_{EXP}(\underline{x}, t)$. At $t = t_0$, each marker was associated with its nearest neighbor $\mathbf{p}_{COMP}(\underline{x}, t_0)$ in the simulated lumen mesh $\mathbf{v}_{CT}(\underline{x}, t_0)$. The mismatch between the time-dependent 3D coordinates $\mathbf{p}_{COMP}(\underline{x}, t)$ computed by SOFA and the ground-truth time-dependent coordinates $\mathbf{p}_{EXP}(\underline{x}, t)$ was quantified as:

$$d(\underline{x}, t) = \|\mathbf{p}_{COMP}(\underline{x}, t) - \mathbf{p}_{EXP}(\underline{x}, t)\|$$

To rule out the effects of initial registration errors, a further metric, the *relative displacement error*, was defined as:

$$\Delta d(\underline{x}, t) = |d(\underline{x}, t) - d(\underline{x}, t_0)|$$

4.4. Results

A video demonstration of the complete framework implementation is available online[2], showing the experimental setup, the XR visualization integrated with real-time sensor data, and the interactive simulation during catheter insertion.

4.4.1. Interactive Performance

The comparison of T_{CPU} (SOFA) vs. T_{PHYS} (*rosbag*) highlighted three phases in the experiments (Table 4.2, Figure 4.5):

1. *Startup latency* – an initial lag by 4.66–6.94 s was observed.
2. *Low-latency phase*, until the catheter tip reached the most tortuous tract of the vein phantom – this phase required 63.83–68.30 s, T_{CPU} progressed slightly more

[2]See the Implementation Video

slowly than T_{PHYS}. The linear regression of T_{CPU} on T_{PHYS} data yielded $T_{CPU} = 1.12T_{PHYS} + 2.10$ ($R^2 = 0.999$).

3. *High-latency phase*, beyond $T_{PHYS} = 56.00$ s – the difference between T_{CPU} and T_{PHYS} markedly increased, with $T_{CPU} = 1.45T_{PHYS} - 18.70$ ($R^2 = 0.993$).

Table 4.2: Cumulative physical time (T_{PHYS}), simulated time (T_{CPU}), and lag at sampled time-points. Ranges indicate variability across three experimental repetitions.

Time-point ID	$T_{\textbf{PHYS}}$ **(s)**	$T_{\textbf{CPU}}$ **(s)**	**Lag (s)**
0	0	4.66–6.94	4.66–6.94
1	8	9.13–13.27	1.13–5.27
2	16	19.43–22.73	3.43–6.73
3	24	28.78–32.34	4.78–8.24
4	32	38.35–41.40	6.35–9.40
5	40	46.97–49.73	6.97–9.73
6	48	56.34–57.93	8.34–9.93
7	56	63.83–68.30	7.83–12.30
8	64	74.04–77.57	10.04–13.57
9	72	82.23–86.75	10.23–14.75
10	80	93.19–97.93	13.19–17.93
11	88	106.74–109.48	18.74–21.48
12	96	121.41–123.67	25.41–27.67

The low variability in the initial lag suggests that startup latency is a fixed delay from system initialization and communication protocols, creating a brief dead zone before data reach the simulation. Lag accumulation during the low- and high-latency phases is mainly due to the increasing computational cost of growing contact interactions. In the low-latency phase, lag grows moderately (slope 1.12 in the T_{CPU} vs. T_{PHYS} regression), as catheter-lumen contact involves few mesh nodes. In the high-latency phase, the expansion of the contact region leads to faster lag accumulation, reflected by the higher slope (1.45; Figure 4.5).

Throughout the experiment, the HL2 rendering FPS remained stable at 35–40 Hz.

4.4.2. Accuracy of Simulated Vessel Deformation

Figure 4.6A shows that the low- and high-latency phases correspond to markedly different vessel deformations, as quantified by the 95th percentile of marker displacement magnitude ($\|\boldsymbol{u}\|_{95}$). During the low-latency phase, $\|\boldsymbol{u}\|_{95}$ remained below 2 mm until the catheter tip reached the most tortuous vessel tract (time-point ID 4), increasing to about 5 mm when the tip was halfway through this tract (time-point ID 8) and the catheter has

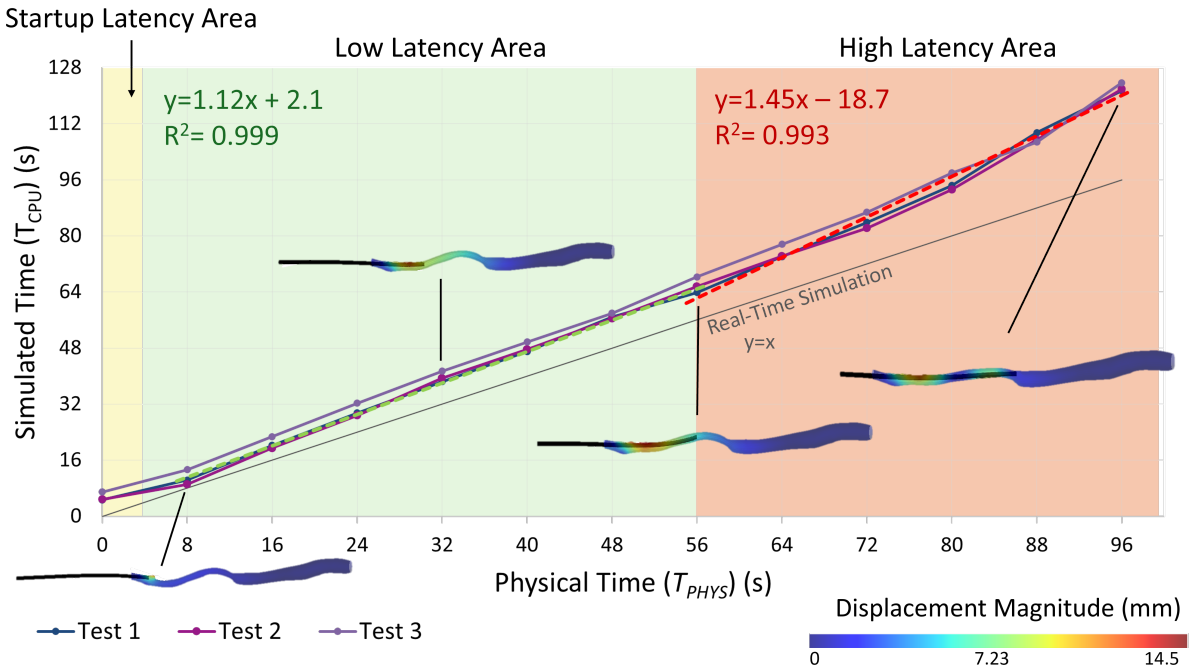


Figure 4.5: Interactive FEM performance: simulated time (T_{CPU}) vs. physical time (T_{PHYS}) for three repeated tests. Three latency phases were identified: startup, low-latency (slope 1.12), and high-latency (slope 1.45). The deformed vein lumen, color-coded by displacement magnitude, is shown at specific time-points.

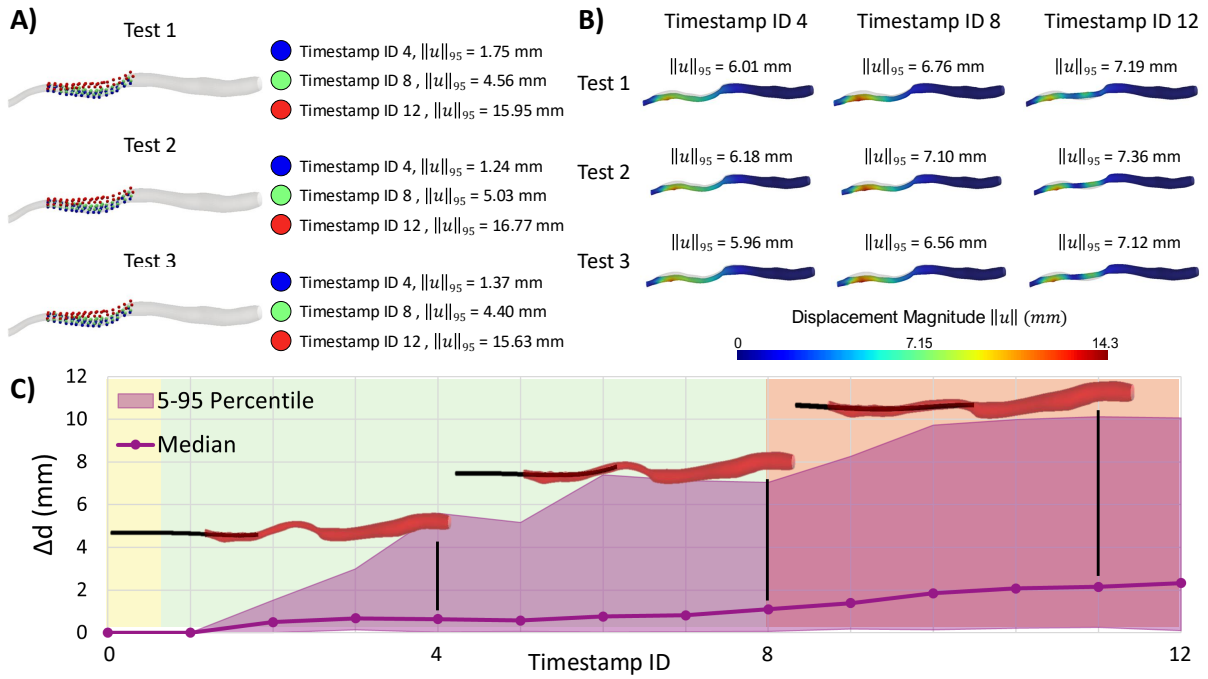


Figure 4.6: **A)** Ground-truth marker configurations ($\mathbf{p}_{EXP}$) overlaid on the undeformed vessel (light gray) at three representative time-points; each is annotated with the 95th percentile of the displacement magnitude ($\|\boldsymbol{u}\|_{95}$). **B)** Corresponding FEM configurations at the same time-points, shown with a displacement colormap and overlaid on the undeformed vessel. The $\|\boldsymbol{u}\|_{95}$ computed over the marker-covered region is reported. **C)** *Relative displacement error* for the three repeated tests, expressed as $\Delta d(\underline{x}_i, t_j)$. Median errors across the 13 time points are shown, with shaded bands indicating the 5th-95th percentile range. Representative FEM vessel-catheter configurations are shown at the three time-points.

not yet straightened the vessel phantom. At the end of the high-latency phase (time-point ID 12), $\|\boldsymbol{u}\|_{95}$ reached nearly 17 mm, corresponding to a clear straightening of the initially tortuous tract.

This straightening behavior was well captured by the FEM model (Figure 4.6B). However, the $\|\boldsymbol{u}\|_{95}$ computed over the lumen region covered by markers underestimated the global deformation, yielding values between 5.96 and 7.36 mm across the three-time points, consistently across tests.

A more reliable comparison is provided by the *relative displacement error* (Figure 4.6C), whose median value remained below 1 mm throughout the early insertion phase, and increased progressively up to 2.33 mm as the catheter advanced through the most tortuous region, where lumen displacements were largest.

The uncertainty in reconstructing $\mathbf{p}_{EXP}(\underline{x}, t)$ arises from stereo calibration re-projection error (0.784 mm) and from the estimation of ${}^{EM}\mathbf{T}_{CAM}$ and ${}^{CT}\mathbf{T}_{EM}$ (0.700 mm and 0.982 mm RMSE, respectively), resulting in a cumulative uncertainty of 2.466 mm. Accordingly, the median *relative displacement error* remained within reconstruction uncertainty throughout insertion. However, as catheter-vessel interaction increased in highly curved regions, the error distribution widened, with the 95^{th} percentile reaching 9.97 mm in the narrowest and most tortuous segment (Figure 4.6C).

4.5. Discussion

4.5.1. Novelty and Technical Contributions

The proposed framework addresses fundamental computational and methodological limitations that have prevented interactive biomechanical simulation from supporting endovascular navigation. While prior work [166] demonstrated interactive catheter-vessel simulation for training applications, extending such approaches to intraoperative monitoring demands more rigorous methodological foundations for modeling surgical interaction.

This work introduces two technical contributions that collectively enable this capability:

Continuum-embedded vessel representation Unlike prior work [166], which modeled vessels as isolated thick-walled conduits with prescribed boundary conditions, we embed the vessel lumen within a deformable continuum (the *BBox*). This design serves two purposes. First, it provides a physically motivated surrogate for the mechanical constraints imposed by surrounding soft tissues, adipose tissue, connective tissue, and muscle,

without requiring patient-specific tissue property maps that are not clinically available from standard imaging. The homogenization strategy represents a pragmatic engineering choice: when detailed heterogeneous tissue models cannot be constructed from clinical data, a compliant continuum with calibrated bulk properties offers a computationally tractable alternative that captures the essential mechanical containment effect. Second, the continuum formulation eliminates the need to infer vessel wall thickness from imaging, which is particularly challenging in venous anatomy.

While this representation is an approximation compared to fully anatomically detailed models that explicitly discretize the vessel wall and surrounding BCs (Section 4.1), it is essential for achieving the computational performance required for interactive execution. The sensitivity analysis (Section 4.3.5) demonstrates that displacement predictions exhibit acceptable stability across physiologically plausible ranges of Young's modulus, confirming the robustness of this modeling assumption under clinically relevant parametric uncertainty.

Future implementations could leverage intraprocedural ultrasound imaging to refine this model. Following CT-ultrasound registration [68, 77], the continuum could be assigned patient-specific mechanical properties derived from non-invasive vessel compliance and distensibility measurements [31], potentially improving prediction accuracy in cases where substantial anatomical heterogeneity is present.

Sensor-driven catheter representation The second contribution addresses a fundamental computational constraint identified through the analysis in Section 4.2.2. Traditional coupled catheter-vessel simulations discretize both structures as deformable finite element models, resulting in a combined system with $(n_c + n_v)$ degrees of freedom where computational cost scales super-linearly with system size and with the number of active contacts. For the catheter-vessel interaction problem, where slender flexible catheters navigate tortuous vessels under geometrically unstable loading conditions, robust contact resolution via LM methods is essential to prevent the numerical instabilities that plague penalty-based approaches. However, the compliance matrix construction and iterative constraint solution (Section 4.2.2) introduce prohibitive computational overhead when both catheter and vessel are mechanically coupled.

By reconstructing the catheter configuration from real-time fiber-optic and electromagnetic sensing and imposing this geometry as a prescribed kinematic boundary condition, the framework eliminates the catheter from the mechanical system entirely. This reduces the simulation to a single deformable body (the vessel) while preserving the numerical stability and zero-penetration guarantees of the LM formulation. The computational benefit

is substantial: factorization cost reduces from $\mathcal{O}((n_c + n_v)^{1.5})$ to $\mathcal{O}(n_v^{1.5})$ and compliance construction from $\mathcal{O}(m \cdot (n_c + n_v))$ to $\mathcal{O}(m \cdot n_v)$.

Critically, this is not merely a design preference but an architectural necessity to achieve interactive performance without sacrificing contact accuracy. The sensor-based strategy transforms an otherwise intractable computational problem into one compatible with interactive execution rates, while simultaneously eliminating the need to tune empirical contact parameters, such as friction coefficients and penalty stiffness, that are difficult to estimate in vivo and are known to be highly application-specific [204] or commonly adjusted heuristically in previous studies [166].

Beyond these technical contributions, this represents, to the best of our knowledge, the first integration of interactive FEM, catheter shape sensing, and XR visualization for endovascular navigation monitoring.

4.5.2. Reliability and Performance

Displacement Accuracy The median *relative displacement error* remained below the cumulative uncertainty of the stereo-based reconstruction pipeline (2.466 mm) throughout catheter insertion. Even the maximum 95$^{\text{th}}$-percentile error (9.97 mm) remained smaller than lumen diameters in both vascular segments (10.19-15.76 mm femoral; 13.38-20.36 mm vena cava). These results support the validity of the continuum-embedded vessel representation for predicting catheter-induced deformations under the experimental conditions tested.

The observed error distribution exhibits spatial heterogeneity: median errors remain low (<1 mm) in regions of modest deformation, while the 95$^{\text{th}}$ percentile increases in the most tortuous segment where contact forces and vessel displacements are largest. The sensitivity analyses (Section 4.3.5) demonstrate that while Young's modulus E influences displacement magnitude (deviations of 0.06-0.56 mm across the tested range), Poisson's ratio ν has negligible impact (<0.05 mm). This parametric robustness is encouraging for clinical translation, as it suggests the model can tolerate reasonable uncertainty in bulk tissue properties without catastrophic degradation in displacement predictions.

Registration Sensitivity Accurate CT-EM registration is critical to the framework's operation. Any offset in the transformation ${}^{CT}\mathbf{T}_{EM}$ uniformly shifts the reconstructed catheter centerline relative to the vessel wall, directly influencing the predicted contact onset location and timing. In the experimental setup, the catheter radius (3.7 mm) and smallest vessel radius (5.095 mm at the femoral insertion cross-section) yield a minimal

radial clearance of 1.395 mm. The measured registration RMSE (0.982 mm) remained below this threshold, ensuring that contact onset occurred at the expected anatomical location rather than being artificially triggered by registration error.

This clearance-to-error ratio serves as a practical metric for assessing framework applicability in new anatomies: registration accuracy must be sufficient relative to the local catheter-vessel clearance to preserve mechanically meaningful contact predictions. In practice, this can be evaluated a priori by inspecting the registered catheter insertion point within the most restrictive vessel cross-section. Beyond the insertion site, the sensitivity to registration error is mitigated by the larger cross-sections along the vessel path, though systematic offsets will still propagate to predicted contact locations.

Interactive Performance and Clinical Context The latency of the proposed system is dominated by the computational delay of the interactive FEM. During the highest-latency phase (32 s of physical time), the FEM accumulated an effective lag of approximately 14 s. While this prevents end-to-end real-time correspondence between physical catheter configuration and simulation state, the update rate remains adequate for interactive clinical use when evaluated within the context of current procedural workflows.

In routine interventional procedures, catheter-vessel interaction is not monitored continuously but is instead assessed intermittently through fluoroscopy. Typical cardiac catheterization procedures last tens of minutes, yet cumulative fluoroscopy time is substantially shorter and delivered in discrete imaging intervals rather than as a continuous stream [8]. Clinicians therefore already operate with temporally sparse and delayed spatial feedback on device-anatomy interaction. Within this clinical reality, a system that provides continuous biomechanical feedback with latency on the order of seconds offers complementary value rather than replacing established imaging modalities.

Furthermore, the current implementation is CPU-based. GPU-accelerated FEM implementations have demonstrated substantial reductions in computation time for soft tissue simulation [41], suggesting a clear pathway for improving interactive performance in future iterations. Additionally, the catheter advancement speed employed here (5 mm/s) was selected based on prior robotic catheter driver studies and Food and Drug Administration (FDA)-cleared clinical systems, but slower advancement would further reduce computational demands during contact-rich navigation phases.

4.6. Limitations and Future Directions

The evaluation presented here was conducted on a single silicone phantom replicating venous anatomy from the femoral vein to the inferior vena cava. While this anatomy exhibits substantial tortuosity and caliber variations that create challenging computational conditions for contact-rich simulation, validation across multiple vessel geometries, including arterial anatomy with different mechanical properties and geometric characteristics, remains necessary to fully characterize the framework's predictive capabilities.

Displacement measurements from external markers are sensitive to phantom wall-thickness variability, which introduces uncertainty in the experimental ground truth. Future validation studies could employ high-resolution dynamic imaging (e.g., biplane fluoroscopy, CBCT) to directly track internal lumen deformation, providing a more robust reference for FEM comparison.

For clinical deployment, the rigid marker-based registration employed here is insufficient. Patient positioning and vessel geometry change between preoperative imaging and the procedure itself. Clinical translation would require deformable CT-ultrasound registration to account for anatomical motion and deformation. The feasibility of such integration is supported by established methods for CT-ultrasound registration [68, 77], which could enable construction of an updated vessel avatar from intraprocedural ultrasound sweeps with known probe poses (e.g., externally tracked or mounted on a robotic arm).

The current analysis focused exclusively on vessel displacement, as contact force predictions are highly sensitive to the mechanical properties assigned to the *BBox*. Without patient-specific tissue property measurements, quantitative force validation would require assumptions that could bias the assessment. Future work incorporating intraprocedural vessel compliance measurements [31] could enable more reliable force prediction and validation.

Finally, this study employed a sensorized catheter instrumented with fiber-optic and electromagnetic tracking. While shape sensing technologies are becoming increasingly available in commercial interventional devices [200, 206], the framework's clinical adoption depends on the availability of such instrumentation in the target procedure. However, the rapid development of fiber-optic sensing for medical applications and the proliferation of electromagnetic tracking in interventional suites suggest that this requirement is becoming less restrictive.

4.7. Final Remarks

This work demonstrates that integrating interactive finite element simulation, real-time catheter shape sensing, and XR visualization can provide continuous monitoring during endovascular catheter procedures. The framework addresses a critical observability gap in current clinical practice: while fluoroscopy reveals catheter position, it does not depict vessel walls, their deformation, or the contact governing vessel displacements. The proposed system makes these invisible mechanical interactions visible to the operator through a XR interface that presents predicted vessel deformation alongside the physical workspace.

5 |

Multisensory Navigation for Percutaneous Cardiac Access

This chapter is based on: Ruozzi, V., Matinfar, S., Vergara, P., Albanesi, A., Dell'Aversana, S., Carugo, S., Buccoliero, G., Navab N. & Votta, E. (2026). "Physics-driven Sonification for Improving Multisensory Needle Guidance in Percutaneous Epicardial Access." – IEEE Access, 2026. This work was conducted in collaboration with the Chair for Computer Aided Medical Procedures & Augmented Reality (CAMP), TU Munich and University of Naples Federico II, Naples.

5.1. Subxiphoid Percutaneous Epicardial Access

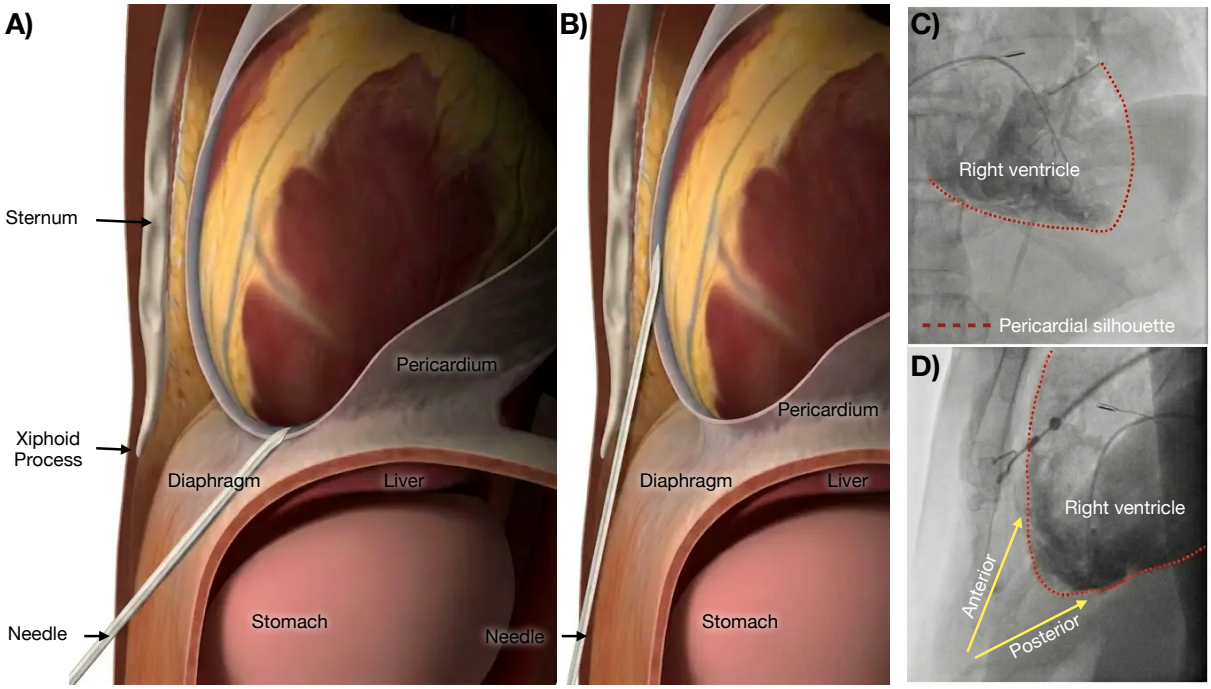


Figure 5.1: Posterior-based (A) and anterior-based (B) approach for epicardial puncture. Needle's insertion under standard fluoroscopic views with pericardial silhouette annotation (C, D). Adapted from [179].

Access to the cardiac anatomical structures for the deployment of therapeutic interventions can be achieved through percutaneous interventions via catheters or needles. Two primary clinical applications of needle-based access are Percutaneous Epicardial Access (PEA) for ablation procedures and pericardiocentesis for emergency drainage.

In electrophysiology, approximately 15% of ventricular tachycardia cases involve subepicardial substrates that cannot be adequately treated through conventional endocardial approaches [194], necessitating direct epicardial access [43]. PEA has emerged as the standard percutaneous technique to address this clinical need. In emergency settings, pericardiocentesis is performed to drain pericardial effusion or hemorrhage. Despite their different therapeutic objectives, both procedures require access to the same anatomical target, the pericardial cavity, and rely on the same access technique, the needle-based subxiphoid approach.

The pericardial cavity is the target anatomical space for both procedures. The cardiac structure is composed of three distinct layers with specialized functions. The pericardium, with an average thickness of 1.2–2.0 mm [22], is a thin membrane that surrounds the heart. It consists of an outer fibrous layer and a dual inner serosal layer, which includes both

parietal and visceral components [22, 202]. The pericardial cavity, situated between the parietal and visceral serosal surfaces, contains 20–30 ml of serous fluid [202]. The visceral pericardium, also known as the epicardium, serves as the outermost layer of the heart wall, followed by the myocardium, the contractile tissue, and the endocardium, the innermost layer.

The needle-based subxiphoid approach, first introduced by Sosa and colleagues in 1996 [193], offers ease of access to the pericardium with reduced costs, rapid patient mobilization, and lower post-interventional morbidity compared to surgical alternatives [43, 194]. The procedure is performed under general anesthesia with controlled ventilation on the beating heart. A rigid Tuohy needle (17/18 gauge, 120–150 mm long) is inserted just below or adjacent to the xiphoid process and advanced toward the pericardial space above the diaphragm [43, 180] (Figure 5.1).

Two distinct approaches are employed based on patient-specific anatomical considerations, especially sternum configuration. While the needle is consistently oriented toward the left shoulder in the frontal plane, the angular direction in the sagittal plane distinguishes the posterior and anterior approaches. The posterior approach employs a sharp insertion angle (steeper than 45° from horizontal) to access the inferior or lateral pericardial surfaces while avoiding coronary vessels (Figure 5.1A). Conversely, the anterior approach utilizes a more shallow angle (0–30° from horizontal) to target the anterior surface of the right ventricle [42] (Figure 5.1B).

The primary difference in procedural complexity between these two clinical contexts lies in the visualization of the target. In emergency settings, the pericardial cavity is typically visibly enlarged due to effusion or hemorrhage, facilitating identification and access. Conversely, in physiological states relevant to electrophysiology procedures, the pericardial cavity contains only 20–30 ml of fluid and exhibits continuous motion with cardiac and respiratory cycles, presenting a considerably more challenging target [193]. Throughout the cardiac cycle, cardiac chamber walls undergo directional displacements ranging from 0.8–2.0 mm in medio-lateral direction, 1.1–2.2 mm in antero-posterior direction, and 1.9–3.1 mm in superior-inferior direction across the four chambers, with the right ventricle specifically exhibiting medio-lateral displacements of 2.0 mm, antero-posterior displacements of 1.1 mm, and superior-inferior displacements of 1.9 mm [76]. Pericardial boundaries exhibit smaller but non-negligible directional shifts of 0.4–0.8 mm [210].

The puncture site is determined solely by palpating bony anatomical landmarks, the xiphoid process and costal margins, without patient-specific trajectory optimization [43]. There is no defined landing zone or anatomical target; the sole requirement is to penetrate

the pericardial cavity without damaging surrounding critical structures. In emergency scenarios, ultrasound guidance is frequently employed, while fluoroscopy-guided needle advancement is the standard for all electrophysiology procedures [43]. During advancement, fluoroscopy provides only two-dimensional projections of the needle and, upon contrast injection, the pericardial silhouette, without visualizing surrounding soft tissues such as the diaphragm, lungs, liver, or coronary vessels [179] (Figure 5.1C,D). Consequently, accurately stopping the needle tip within the thin pericardial space without perforating the underlying myocardium is exceptionally challenging, particularly given cardiac and diaphragmatic motion [26, 180]. As the visual feedback provided by imaging systems is limited and constrained (Section 2.1.1), tactile feedback is very important: as the fibrous pericardium is reached, operators rely on tactile feedback, perceiving a characteristic release of resistance as the needle penetrates the parietal pericardium [26]; cardiac motion transmitted through the needle may also be appreciated at this point [180]. It is crucial to adequately interpret this feedback to promptly stop the advancement of the needle, but its correct interpretation requires a considerable expertise level of the operator.

Following successful pericardial puncture, a guidewire is inserted into the pericardial cavity, enabling advancement of a hollow sheath or a syringe, in case of pericardiocentesis. In electrophysiology, flexible catheters are subsequently introduced through the sheath to access the epicardium for cardiac mapping and energy delivery (radiofrequency or cryoablation) to ablate arrhythmogenic tissue [42].

Despite widespread adoption in electrophysiology, with utilization in 23–30% of ventricular tachycardia ablations [14], the procedure remains technically demanding and carries substantial risk. This is due to the current approach for determining insertion site and needle orientation, and to guidance conditions that rely heavily on the operator's experience. The risk of major PEA-related complications ranges from 6% to 15% [134, 171]. In a cohort of 271 epicardial ablation procedures, 34 major complications (12.5%) were observed, with 23 cases (8.5%) directly attributable to the epicardial puncture itself [134]. The most common severe complications include pericardial bleeding from inadvertent right ventricle laceration and coronary vascular injury, with unintentional right ventricle puncture reported in 4.5–17% of attempted procedures [171, 179]. Additional documented complications include perforation of abdominal organs (liver, colon, stomach), damage to epigastric or internal mammary arteries, lung puncture, and injury to phrenic nerves [134, 171].

This elevated complication rate is primarily attributable to the absence of patient-specific preprocedural planning and the limitations of guidance modalities: imaging systems cannot adequately visualize critical soft-tissue structures along the needle trajectory,

while attenuated and subjective tactile feedback cannot reliably inform on time-critical events [43, 179].

So far, no methodology exists allowing real-time accurate guidance in highly dynamic settings like PEA.

5.1.1. Proposed Framework

In this Chapter, a novel XR-based multisensory navigation system designed to support situational awareness during the critical needle landing phase on beating cardiac structures in PEA is proposed and evaluated. The novelty lies in addressing the challenge of needle guidance in the presence of moving targets.

The system integrates visual and auditory feedback: an OST-HMD displays 4DCT-based time-resolved anatomy, static planned navigational cues, and a real-time needle avatar to establish spatial context, while physics-driven sonification delivers detailed, real-time guidance during navigation toward the beating heart. The primary objective is to provide precise guidance, ensuring that the needle reaches the pericardial membrane without penetrating the myocardium, which represents the complication with greatest clinical severity.

The system was developed in close collaboration with medical experts to align with clinical requirements and is intended to complement rather than replace standard fluoroscopy guidance, with particular benefit as the needle approaches the pericardial membrane. Beyond this stage, conventional fluoroscopic views remain indispensable for subsequent procedural steps, such as final needle positioning and tool deployment within the pericardial cavity.

To evaluate this approach, the framework was implemented in an experimental set-up and a phantom study was conducted with 12 cardiologists.

5.2. Methods

The proposed method is illustrated in Figure 5.2 as a modular workflow integrating three main components: (1) patient-specific anatomical reconstruction and surgical planning; (2) real-time tool navigation and computation of navigation parameters; and (3) multisensory feedback generation, with particular emphasis on physics-driven auditory cues. These components operate in a closed-loop manner to ensure temporal and spatial coherence between tool motion and user guidance. The multisensory feedback is delivered to the user through an XR system comprising two display modalities: a visual inter-

face implemented via an OST-HMD and an acoustic interface provided through external speakers.

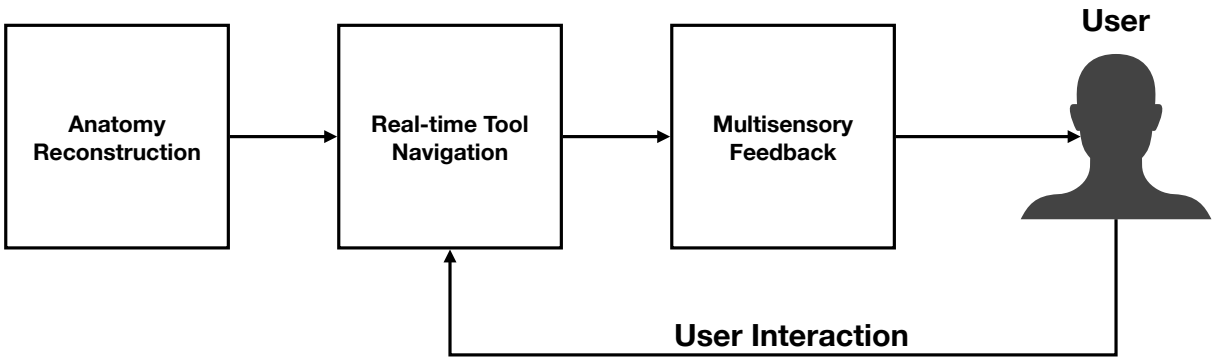


Figure 5.2: High-level schematic of the proposed workflow.

The detailed implementation of each component for PEA is shown in Figure 5.3 and described in Sections 5.2.1, 5.2.2, and 5.2.3. As the core contribution lies in the methodology and evaluation of the multisensory guidance system (Figure 5.3C), Section 5.2.4 focuses on the experimental setup and the design and execution of the user study conducted to assess this component.

5.2.1. Dynamic Anatomy Reconstruction and Trajectory Planning for PEA

A preoperative ECG-gated 4DCTA dataset ($0.785 \times 0.785 \times 0.600$ mm spatial resolution, 20 frames per cardiac cycle at 5% temporal intervals) was analyzed (Figure 5.3A).

Image processing was conducted in 3D Slicer [55]. An automatic segmentation of skin surface and of all main anatomical thoracic structures was obtained on the reference End-Diastolic Frame (EDF) CT volume through TotalSegmentator algorithm embedded within 3D Slicer [220]. The cardiac muscle (i.e., myocardium) was automatically segmented in each frame. The external pericardial membrane (i.e., pericardium), which presents greater segmentation complexity, was manually segmented in the EDF by delineating pericardial boundaries with the surface cut tool.

Segmentations for the other frames were obtained through Deformable Image Registration (DIR) [103], by registering each frame to the EDF used as the fixed reference. The resulting sequence of ${}^{mov}T_{fix}$ transforms was applied to the pericardium segmentation mask, enabling dynamic pericardial profile segmentation across all time frames. The surfaces of the myocardium and the pericardium were reconstructed from the segmentation masks, and their distance (i.e. inter-parietal distance) was computed for each of the 20

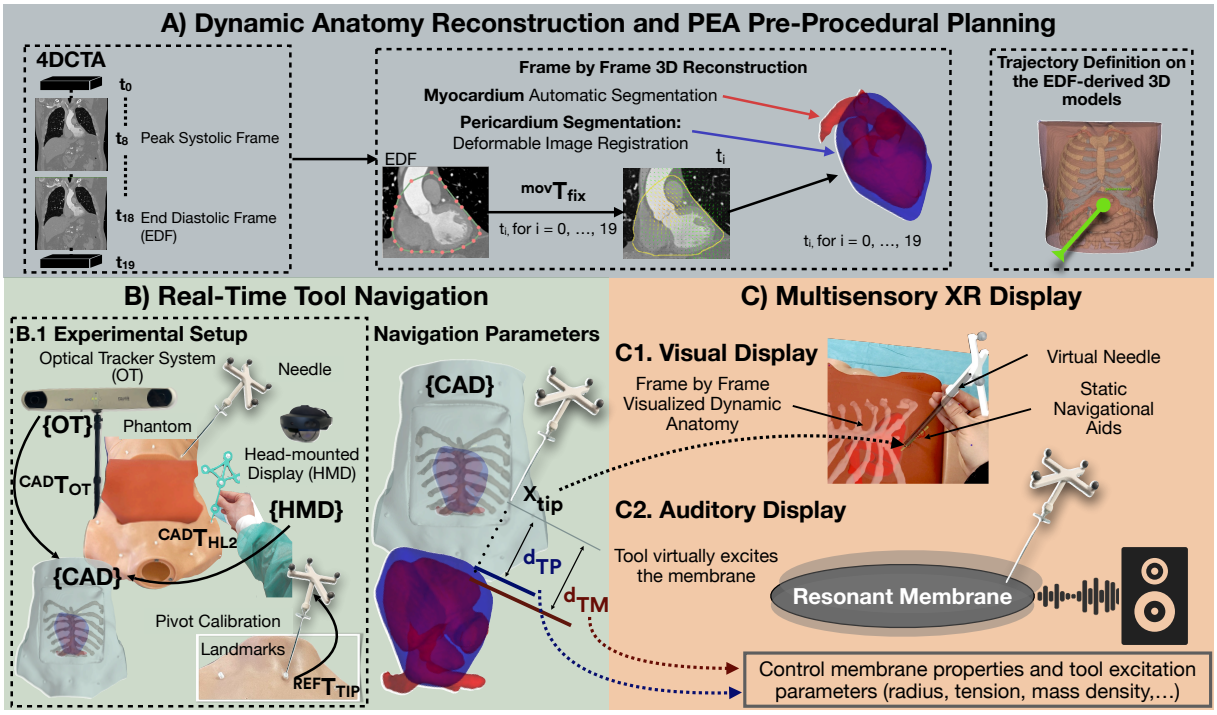


Figure 5.3: Overview of the XR-based multisensory surgical navigation system developed for PEA. (A) ECG-gated 4D Computed Tomography Angiography (4DCTA) frames over one cardiac cycle were analyzed. The myocardium and the external pericardial membrane (i.e., pericardium) were segmented, and their three-dimensional models reconstructed. An *ad hoc* planning algorithm was then applied to compute risk-optimized needle trajectories, avoiding surrounding critical organs. (B) Real-time computation of navigation parameters, defined as the Euclidean distance between the needle tip (x_{tip}) and the moving pericardium (d_{TP}) and myocardium (d_{TM}). A navigation setup for real-time needle tracking was implemented. Registration steps were required to coherently track the needle tip (x_{tip}) relative to the CT-derived phantom model ({CAD}), enabling the computation of navigation parameters. (C) The XR-based multisensory display integrated modular visual and auditory components with input data prepared as detailed in (B), to provide accurate and dynamic tool guidance.

frames. The validation of this registration-based approach is available in Appendix B.

By analyzing the temporal variation of this distance throughout the cardiac cycle, safer needle landing zones, corresponding to regions of maximum inter-parietal distance, were identified. This information, combined with segmentation of critical organs (e.g., coronary, epigastric and mammary arteries, lungs, and liver) in the EDF frame, was used to initialize the preoperative planning platform.

Planning was performed considering the EDF as the worst-case scenario, since at this point in the cardiac cycle the myocardium is expected to reach its maximum volume, resulting in the minimum inter-parietal distance. An *ad hoc* algorithm, shown in Figure 5.4, was implemented to compute optimal trajectories through three consecutive filters: (1) the *collision filter*: a ray-casting algorithm discarded any trajectory intersecting critical organs; (2) the *length filter*: the total path length was compared against the needle length; and (3) the *distance filter*: trajectories were refined to ensure a minimum clearance of 5 mm from all segmented sensitive structures, in accordance with clinical recommendations for epicardial punctures [15, 104].

5.2.2. Real-Time Tool Navigation

The proposed physics-driven sonification method (Section 5.2.3) relies on real-time needle tip tracking and the computation of navigation parameters. Specifically, the system continuously computes the Euclidean distances along the needle axis between the needle tip ($\mathbf{x}_{tip}$) and the moving pericardium and myocardium, denoted as d_{TP} and d_{TM}, respectively (Figure 5.3B).

5.2.3. Multisensory Extended Reality Surgical Navigation Display

The proposed XR surgical navigation system integrates complementary visual and auditory feedback modules to support precise tool guidance during pericardial access procedures. Both modules operate in real-time using the same input data stream, ensuring spatial and temporal coherence while maintaining full independence from each other.

In the following sections, the visualization and sonification modules, together with their respective data-processing pipelines, are described.

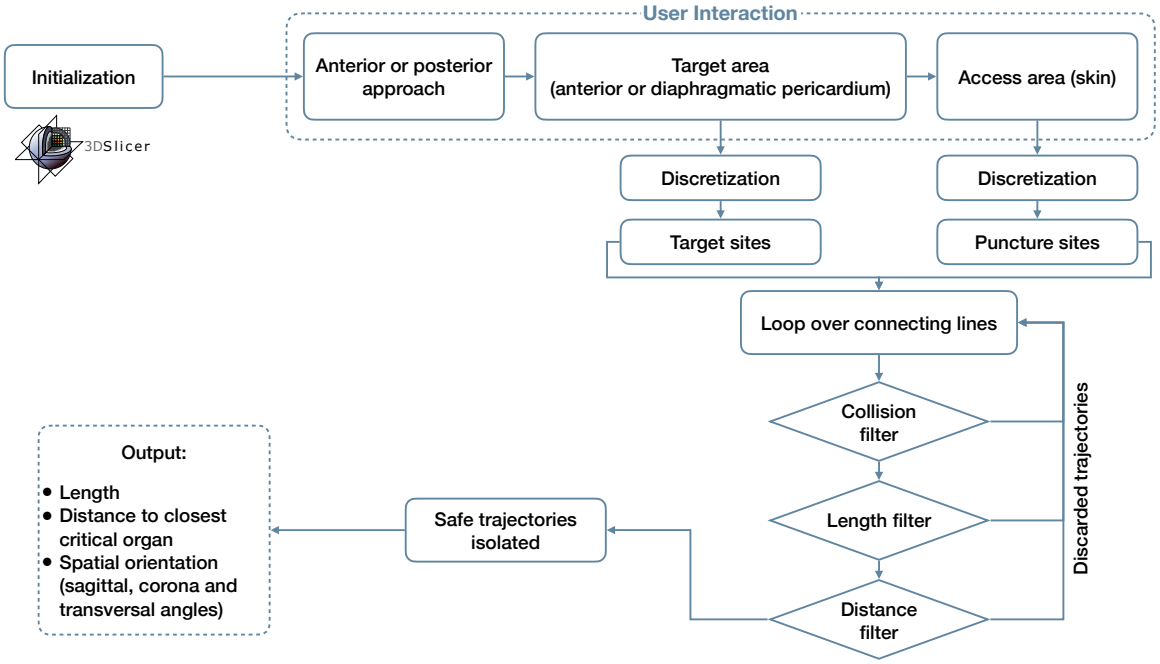


Figure 5.4: Schematic illustration of the *ad hoc* algorithm's workflow to isolate safe insertion trajectories for patient-specific epicardial punctures. The algorithm begins with user selection of either an anterior or posterior approach, restricting the search to the associated pericardial surface. Three primary segmentations are automatically classified: critical structures to avoid during needle insertion, skin access surface, and target pericardial area (anterior or diaphragmatic portion). An oriented bounding box tree locator is initialized for collision detection. The user places markers on the skin segmentation to define a custom access region. Access and target areas are discretized with a resolution of $\frac{1\,\text{marker}}{0.5\,\text{cm}^2}$, distributing potential puncture and target sites accordingly. For each target site, the algorithm loops through all puncture sites to isolate risk-free trajectories satisfying the three filters. All trajectories are displayed with key measurements: length, angular inclination, and minimum distance from critical organs.

Visualization

The XR visual interface was built in Unity3D (Unity Technologies, v2020.3.1 LTS) and deployed to the OST-HMD device. The 3D reconstruction of the patient-specific anatomy for 20 distinct time frames sampled over one full cardiac cycle (Section 5.2.1) was visualized. These reconstructed frames sequentially animated in a continuous loop during runtime, providing a dynamic reliable replica of the beating cardiac structures. Static visual cues indicated the planned trajectory: a crosshair marks the access point, while a fixed cylinder aids needle alignment. An array of rings further enhances alignment[33]. A virtual model of the needle was also added to the scene. Its pose was updated by the real-time $\mathbf{x}_{tip}$, received via TCP.

Sonification

For sonification to be reliable and effective, it must fulfill two key criteria: (a) **Informational Clarity:** The auditory representation must convey critical insights necessary for the safe execution of the task in a clearly perceptible manner. (b) **Intuitiveness and Usability:** The sonification must be easily comprehensible, allowing users to quickly learn and apply it in high-stakes surgical contexts. It should impose minimal cognitive load, ensuring that it does not interfere with the primary task.

Informational Clarity – State-Based Sonification To ensure safe puncture execution and accurate needle positioning within the epicardial space, the auditory representation must clearly convey critical spatial information relevant to the task. In particular, it should precisely indicate the tool-tip's proximity to both the pericardial membrane and the myocardium. Furthermore, it must signal membrane perforation, enabling the user to halt needle advancement and achieve correct positioning without contacting the myocardium.

To achieve this, four distinct **sonification states** were defined, presented in Figure 5.5, that must be clearly conveyed: *(1) Outer Pericardial Zone,* the region outside the pericardium, where the user requires guidance to navigate toward the critical puncture area, encompassing key procedural stages. *(2) Pre-Puncture Zone,* the area immediately adjacent to the pericardial membrane, indicating that the user is nearing, but has not yet reached, the puncture site. *(3) Safe Puncture Zone,* the target region where the pericardial membrane has been successfully reached, and the puncture is safely performed. *(4) Myocardium (Danger Zone),* a high-risk area that must be strictly avoided to prevent complications.

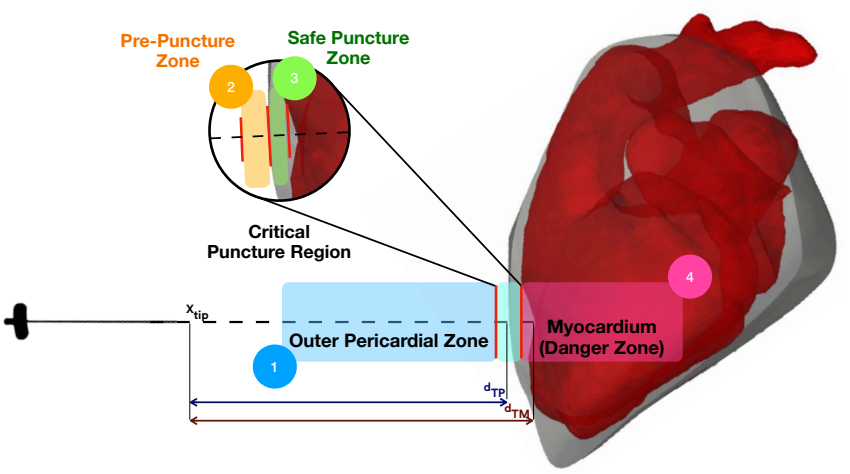


Figure 5.5: Conceptual representation of anatomical regions in PEA for state-based sonification. Transitions are defined based on the Euclidean distance computed along the needle axis from $\mathbf{x}_{tip}$ to both the moving pericardium (d_{TP}) and myocardium (d_{TM}). The figure is for illustration purposes only; size and order of magnitude are not to scale.

Intuitiveness and Usability – Physical Modeling Synthesis The complexity of this sonification scenario, involving multivariate and multistate signals, poses a challenge in delivering auditory cues that are both clearly conveyed and easily learnable, while ensuring they do not interfere with task execution.

A straightforward one-to-one mapping approach, in which each data dimension is directly assigned to a single auditory parameter (e.g., mapping distance X to sound frequency), is simple to design but not necessarily optimal from an interaction perspective [79].

Conversely, a one-to-many mapping approach, in which each data dimension is represented by multiple auditory features, can enhance clarity and perceptual salience [79]. However, improving usability requires sonification to provide causally driven sounds, namely, behaviorally familiar and physically meaningful auditory cues that enhance intuitiveness [147, 208]. Recent studies have demonstrated the effectiveness of embedding causal relationships in medical sonification [140, 182].

To address the challenges posed by heart dynamics and dual-layer distance, physical modeling synthesis was employed, which enables the definition of virtual objects that emulate realistic acoustic behaviors [40, 191].

2D Circular Membrane Model for Sonification A 2D circular membrane model was employed to represent the dynamic behavior of the heart structures. A circular membrane is a thin, flexible surface that vibrates when excited, similar to a drumhead. Its

oscillatory behavior is governed by well-defined vibrational modes, which simultaneously determine multiple auditory parameters such as frequencies, amplitudes, and damping characteristics, thereby inherently realizing a one-to-many mapping. The membrane's response is influenced by its radius, material properties such as surface density, tension, and damping factors, which control how energy dissipates over time. A subset of these factors was used for mapping, with Section5.2.3 and Table5.1 providing detailed parameterization of these factors. The membrane's oscillations combine to produce an output signal $s(t)$ derived from membrane vibrations that can be described as a sum of its vibrational modes:

$$s(t) = \sum_{m,n} A_{m,n} e^{-\sigma_{m,n} t} \cos(2\pi f_{m,n} t + \phi_{m,n}), \tag{5.1}$$

where $A_{m,n}$ is the amplitude of mode (m, n), determined by the applied force; $f_{m,n}$ are the natural frequencies determined by the membrane's physical parameters, including its radius, tension, and mass density, and (m, n) are the mode indices, defining spatial vibration patterns; $e^{-\sigma_{m,n} t}$ represents damping, and is defined as: $e^{-(\beta+\alpha \cdot f_{m,n}^n) \cdot t}$, where β and α are constant and frequency-dependent loss coefficients; $\phi_{m,n}$ is the initial phase of each mode.

For reproducibility, only the synthesis parameters explicitly defined in Section 5.2.3 and Table 5.1 were modulated during runtime, while all other membrane parameters were kept constant. The excitation force F was held constant across all states, and the initial phase terms $\phi_{m,n}$ were fixed at implementation defaults.

This formulation allows for realistic temporal decay of each mode, enhancing the acoustic representation of cardiac structures behavior.

Data Mapping to Synthesis Parameters The previously defined sonification states can be formally classified based on two input features, d_{TP} and d_{TM} using two threshold conditions, as follow:

- State 1: $(d_{TP} > 5.0 \text{ mm}) \wedge (d_{TM} > 2.0 \text{ mm})$
- State 2: $(d_{TP} \leq 5.0 \text{ mm}) \wedge (d_{TM} > 2.0 \text{ mm})$
- State 3: $(d_{TP} \leq 0.0 \text{ mm}) \wedge (d_{TM} > 2.0 \text{ mm})$
- State 4: $(d_{TM} \leq 2.0 \text{ mm})$

State classification was performed using the original (non-normalized) distance values d_{TP}

and d_{TM}.

Each sonification state modulates one or more of the membrane's parameters, influencing its dynamic response and resulting sound characteristics. The affected parameters are: R (membrane radius, affecting both fundamental and modal frequencies; larger $R \Rightarrow$ lower $f_{m,n}$, thus lowering perceived pitch), β (constant loss coefficient, controlling global damping; higher $\beta \Rightarrow$ shorter sustain via $\sigma_{m,n}$), α (frequency-dependent loss; higher $\alpha \Rightarrow$ stronger high-frequency damping, shaping spectral decay), F (excitation force, determining mode amplitudes $A_{m,n}$), and Δt (timing of force application, shaping rhythmic structure). This mapping dynamically modulates the membrane's acoustic properties, pitch, damping, spectral content, and rhythmic excitation, based on input variations.

The implementation modulates only the subset of parameters explicitly specified in Table 5.1. In practice, membrane radius R was operationalized through the corresponding fundamental frequency f_0, which is reported in the table. The excitation force F was kept constant across all states, while β, α, and Δt were dynamically updated according to the current state and normalized input values. Parameters not explicitly updated in a given state retained their previously assigned values within that state, ensuring continuity across state transitions.

The two distances (d_{TP} and d_{TM}) were received via Open Sound Control (OSC) in millimeters and normalized using the same linear clamping and scaling function. Each distance $d \in \{d_{TP}, d_{TM}\}$ was mapped to a normalized variable $\hat{d} \in [0, 1]$ as

$$\hat{d} = \frac{\text{clip}(d, d^{\min}, d^{\max}) - d^{\min}}{d^{\max} - d^{\min}},$$

where $\text{clip}(x, a, b) = \min(\max(x, a), b)$. The bounds were set to $d_{TP} \in [1, 60]$ mm and $d_{TM} \in [1, 30]$ mm.

The normalized variables $\hat{d}_{TP}$ and $\hat{d}_{TM}$ directly determine the parameter ranges and interpolation intervals specified in Table 5.1.

State assignment was evaluated hierarchically in the following priority order: State 4 $\rightarrow$ State 3 $\rightarrow$ State 2 $\rightarrow$ State 1, to avoid ambiguity when threshold conditions overlap. Within each state, parameter values were updated by linear interpolation over the corresponding normalized input intervals specified in Table 5.1.

For consistency with Table 5.1, normalized variables are expressed in descending intervals from 1 (far) to 0 (contact).

These mappings are designed to ensure that the resulting sounds fall not only within the

Table 5.1: State-Dependent Parameter Modulation. The normalized input variables $\hat{d}_{TP}$ and $\hat{d}_{TM}$ define the interpolation intervals within each state. Membrane pitch is reported as fundamental frequency f_0, which operationalizes membrane radius R. Temporal excitation intervals Δt are given in milliseconds.

State	$\mathbf{f_0}$ (Hz)	β	α	$\mathbf{\Delta t}$ (ms)
1	100	2	10	$d_{TP} \in [1, 0.5)$ $\rightarrow \Delta t \in [500, 271]$
2	100	$d_{TM} \in [1, 0.5)$ $\rightarrow \beta \in [2, 1.06]$	$d_{TM} \in [1, 0.5)$ $\rightarrow \alpha \in [10, 5.075]$	$d_{TP} \in [0.5, 0)$ $\rightarrow \Delta t \in [270, 40]$
3	400	$d_{TM} \in [0.5, 0]$ $\rightarrow \beta \in [1.05, 0.1]$	$d_{TM} \in [0.5, 0]$ $\rightarrow \alpha \in [5.075, 0.15]$	40
4	1000	0.1	0.15	40

range of optimal auditory perception but also within a perceptual comfort zone, where differences are clear, easily distinguishable, and not fatiguing to the listener [235]. The parameter values in Table 5.1 were selected to create perceptually distinct sonic spaces. Fundamental frequencies of 100, 400, and 1000 Hz are separated by more than one auditory critical bandwidth ($\approx$ 100 Hz at 400 Hz, $\approx$ 160 Hz at 1000 Hz), ensuring robust pitch discrimination [235]. The damping parameters α (10–0.15) and β (2–0.1) control the exponential decay of the oscillatory response, shaping loudness envelopes with differences comfortably above detection thresholds of $\sim$ 1 dB and thereby yielding clearly distinct decay behaviors [148]. The temporal parameter Δt spans 500–40 ms, covering perceptual tempo categories from slow ($>$ 500 ms) to fast ($<$ 100 ms), with differences well beyond the $\sim$ 20–30 ms temporal resolution limit [235]. Altogether, this parametrization provides a systematic basis for mapping physical dynamics to perceptually effective auditory cues.

With this approach, perceptual sound attributes are not prescriptively selected, as in conventional parameter-mapping methods; instead, they emerge from the resonator dynamics, shaped jointly by the initial model parameters and their anatomy-inspired, state-dependent, and interaction-driven modulations.

Cognitive Auditory Mapping and Learned Associations This mapping ensures that variations in input data dynamically modulate the membrane's acoustic properties, influencing pitch, damping, spectral content, and rhythmic excitation patterns. The rhythmic excitation follows an echolocation-inspired mechanism, analogous to that used in parking assistance systems, serving as a fundamental cue for distance perception. Energy

dissipation (regulated by α and β) emulates the myocardium's dynamic damping behavior during diastole and systole, where varying proximity to the pericardial membrane modulates resonance and provides an intuitive, perceptually salient auditory cue. In addition, modulating the membrane radius (R) induces pitch variation, reinforcing perceptual distinctions between interaction zones. When the tool reaches the myocardium threshold, the radius, damping coefficients (α, β), and spectral characteristics shift abruptly, producing a clear and immediate auditory warning of critical proximity.

It is worth noting that the resulting auditory feedback emerges from a model that reflects anatomical context and user interaction, preserving causal coherence between interaction, anatomical state, and sound. This enables systematic and extensible sonification in complex, time-varying surgical scenarios such as needle guidance under cardiac motion. In contrast, conventional approaches predominantly rely on explicit parameter-mapping strategies that directly map low-dimensional signals to sound parameters (e.g., pitch or rate), often requiring manual signal segmentation and heuristic tuning as interaction complexity increases.

A video illustrating the auditory mapping framework and demonstrating the resulting sounds is available online[1].

5.2.4. System Evaluation: The Phantom Study

To evaluate the proposed XR-based multisensory surgical navigation system in a controlled yet clinically meaningful environment, a phantom study was conducted. This section describes the experimental setup used to simulate PEA and the design of the user study.

Experimental Setup

The experimental setup (Figure 5.3B.1) consisted of: (i) a phantom with a rigid sternum, ribs, and a puncturable foam chest patch; (ii) an optical tracker (NDI Polaris Spectra); (iii) a 18 gauge Tuohy needle (150 mm long) with a custom three-dimensional (3D)-printed passive marker array; and (iv) an OST-HMD device. A CT scan of the phantom was acquired to generate its 3D virtual reconstruction. Since the physical phantom did not include cardiac structures, its reconstruction was added to the virtual scene described in Section 5.2.3 and registered to the animated cardiac anatomy by aligning the sternum structures.

[1]See the Sonification Design Video

In this XR framework, three Frame of Reference (FoR) were involved: the optical tracker (i.e. {OT}), the phantom anatomy reconstructed from the CT scan (i.e. {CAD}), and the OST-HMD FoR (i.e. {HMD}). To enable real-time computation of the navigation parameters such as the needle tip's relative distance to the anatomy, and to ensure the coherence of the information among all the FoRs, calibration and registration procedures were required. First, the transformation ${}^{REF}T_{TIP}$ was computed to correctly track the needle's tip with respect to the optical tracker FoR ({OT}), achieving a pivot calibration RMSE of 1.09 mm. Considering the {CAD} as the global FoR, two marker-based registration procedures (Section 2.1.3 with the {OT} and the {HMD} were performed. Eight markers were designed to accommodate the needle's tip in a unique manner, and were attached to the phantom. Their precise location within the {CAD} were reconstructed from the CT scan. Markers' positions were acquired both in the {OT} space with the tracked needle, and in the {HMD} space using another tool tracked directly by the OST-HMD [90]. The accuracy of the marker-based registration was evaluated by computing the RMSE. The transformation from {OT} to {CAD} space (i.e. ${}^{CAD}T_{OT}$) yielded an RMSE of 1.2 mm, whereas the registration of {HMD} with {CAD} space (i.e. ${}^{CAD}T_{HMD}$) resulted in an RMSE of 2.2 mm.

This procedure ensured a correct tracking of the needle's tip in the {CAD} space (i.e. x_{tip}). This information was streamed in real-time via the Transmission Control Protocol (TCP) to OST-HMD. Moreover, d_{TP} and d_{TM} were computed and streamed to the sonification module via the OSC protocol [224].

The User Study

The user study was co-designed with medical experts to ensure clinical relevance and alignment with task requirements. The goal was to assess whether the proposed multisensory guidance improves safe needle navigation toward the pericardial membrane compared to visual-only guidance. The system is intended to support the operator specifically when the needle approaches the pericardial membrane. Guidance success was defined as reaching the pericardial membrane while avoiding contact with the underlying beating myocardium. Participants were asked to perform a series of punctures by following predefined trajectories. As the primary focus of this study was dynamic target navigation, the initial needle-to-path alignment was implemented following the method presented in [140] and was not part of the evaluation. Each trial started with a start command once correct alignment was established and ended when the participant verbally confirmed completion by saying *"STOP"*, indicating user's confidence in the needle position at which they would have proceeded with the next procedural steps.

Relying the experimental navigation system (Section 5.2.4), sonification was implemented via MAX/MSP[2] for real-time audio processing and Modalys[3] for physical modeling synthesis (Section 5.2.3).

Six trajectories were generated using the planning algorithm (Section 5.2.1), targeting three distinct points (1, 2, 3) on the pericardium (Figure 5.6). A medical expert validated the trajectories for consistent difficulty and clinical relevance, and they were presented in uniformly randomized order to the participants. The assignment of trajectories was balanced across modalities to ensure comparable task difficulty between V and Multisensory (MS) conditions.

The target population for the study consisted of cardiologists and electrophysiologists. They had no prior experience with the system: they first received a brief introduction, followed by a standardized explanation of the study procedure ($\backsim$ 5 min). They then performed a self-exploration of the system, completing three punctures with visual feedback and three with multisensory feedback ($\backsim$ 5 min). This familiarization phase was designed to provide comparable initial exposure to both modalities and reduce potential learning effects during the evaluated trials.

Each participant performed ten trials: five punctures under visual-only (V) guidance and five under MS guidance. The modality order was counterbalanced across participants (V-MS vs. MS-V) to mitigate potential order and learning effects, with a short break of approximately 5 minutes between blocks to reduce carry-over effects. Within each modality block, trial order was randomized to further minimize sequence effects.

Execution time was recorded from the start command to the participant's verbal confirmation of completion. Throughout each puncture, the needle tip's intersections with the dynamically moving cardiac structures were continuously monitored with a ray casting-based collision detection algorithm. Specifically, potential intersections with the pericardium and myocardium were recorded during the whole task within the 20-frame cardiac cycle animation. Additionally, the final needle tip position at the moment of completion was logged. After each condition, participants completed the NASA Task Load Index (NASA-TLX) workload questionnaire, rated perceived difficulty of each modality, and provided subjective feedback.

This experimental design, combining counterbalanced modality order, randomized trial sequences within each block, and prior familiarization, was designed to mitigate potential learning and carry-over effects between conditions.

[2]https://cycling74.com/products/max
[3]https://support.ircam.fr/docs/Modalys/current/

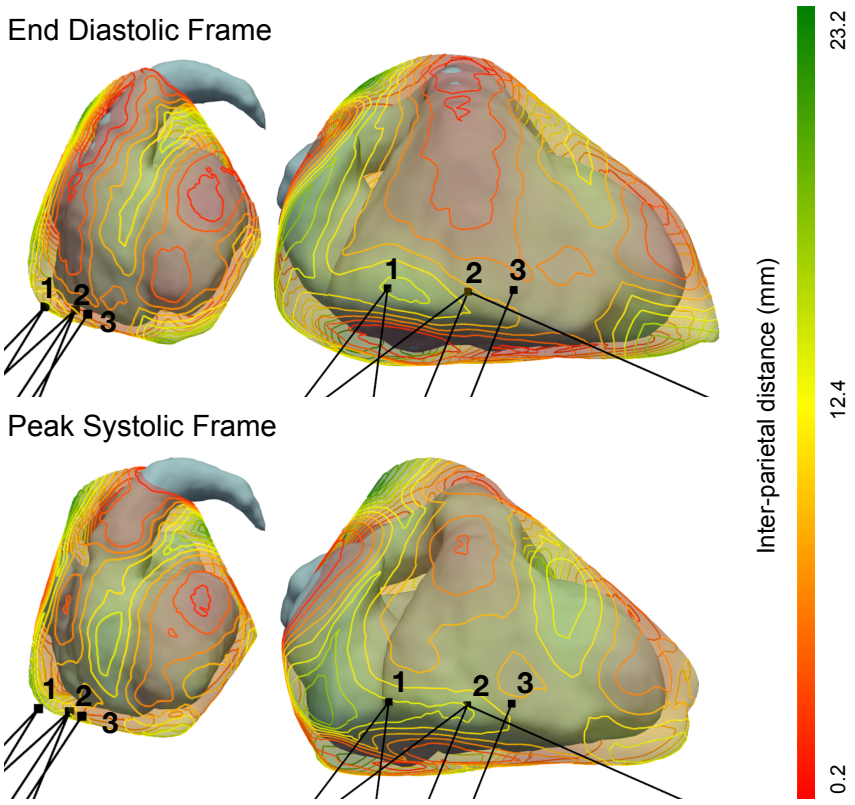


Figure 5.6: Planned trajectories for the phantom study, shown at End-Diastolic and Peak-Systolic phases of the cardiac cycle. Six trajectories were generated toward three target points (1,2,3) on the pericardium, following the strict requirements of PEA (Section 5.2.1). The myocardium and the initial tract of the pulmonary artery are shown in gray. The pericardium is color-coded to represent the inter-parietal distance between the myocardium and pericardium. This distance varies both spatially and temporally: it increases during Peak Systole, whereas End Diastole represents the most critical phase, with the myocardium at maximum volume and a higher risk of myocardial injury.

5.3. Results

The experimental data collected as described in Section 5.2.4 were analyzed investigating several aspects relevant for intraoperative guidance in PEA. A video showing the complete framework implementation during the user study is available online[4].

First, navigation task was assessed by determining whether participants were able to safely reach the pericardium surface while avoiding myocardial contact.

Subsequently, the final needle tip position at the moment of the *"STOP"* command was examined, considering only safe (successful) trials. Since the system was designed to guide the operator in accurately reaching the pericardium, the needle tip position at the point when participants felt confident to stop was considered relevant, as it represents the starting point of subsequent procedural steps under fluoroscopic guidance (as outlined in the Introduction). At this stage, the needle was fixed within the phantom's foam chest patch; thus, the recorded tip position reflected the operator's perceived alignment with the intended target. In contrast, the final needle tip orientation was not evaluated because the study focused on dynamic target navigation rather than trajectory-following accuracy, with initial needle-to-path alignment handled as in [138]. Furthermore, orientation at the end of advancement has limited clinical relevance once safe pericardial entry is achieved, as it does not impose strict constraints on subsequent catheter deployment. This final position was therefore used to assess the accuracy of needle placement according to two aspects: (1) accuracy of the pericardial access: once the sac was reached, how promptly did the user perceive the moving membrane and stop the needle, while maintaining minimal penetration depth to ensure tip placement suitable for subsequent catheter deployment? This was quantified as the minimum distance from the final needle tip position to the moving membrane; and (2) the spatial accuracy: where was the final needle tip located relative to the predefined anatomical target for each trajectory? This was quantified as the Euclidean distance between the final needle tip and the planned target point.

This section is structured reporting the results of these analysis.

The study included 12 participants (age range: 25–47 years; 10 male, 2 female), comprising 2 cardiologists, 3 electrophysiologists, and 7 medical residents with heterogeneous experience in PEA.

The participants exhibited diverse levels of procedural experience: 3 performed PEA weekly, 2 had completed it fewer than five times, 3 had assisted in procedures, and 4 had no prior experience. Most reported limited musical practice (either passive listening or

[4]See the User Study Video

rare instrument use) and variable exposure to OST-HMD technology, ranging from no prior use to occasional experimentation.

5.3.1. Navigation Task Outcome Analysis

Each trial was evaluated by classifying the navigation outcome into three categories based on the intersection data recorded during navigation up to the user's verbal *"STOP"* confirmation. Each participant performed ten trials (five under V guidance and five under MS guidance):

- *Successful Completion*: reaching the pericardial membrane without myocardial contact;
- *Missed Target*: failure to reach the pericardial sac during the navigation;
- *Critical Failure*: unintended contact with the myocardium.

The study comprised 120 evaluated trials in total (12 participants $\times$ 10 trials), including 60 trials under V guidance and 60 under MS guidance. Task outcome rates were computed over all trials within each modality (n = 60 per condition).

For both modalities, the distribution of task outcome rates was computed according to the three categories defined above and performed Chi-square tests to evaluate statistical significance between the groups.

Across all participants, the *Successful Completion* rate was higher under MS guidance (83.3%, 50/60) than with V guidance (55.0%, 33/60). Conversely, the incidence of *Missed Targets* was lower under MS (13.3%, 8/60) compared to V (36.7%, 22/60). Similarly, the rate of *Critical Failures* decreased with MS (3.3%, 2/60) relative to V (8.3%, 5/60). Statistical analysis confirmed a significant effect of modality on task outcome ($\chi^2 = 11.30, p < 0.01$).

Expertise Level Subgroups: An independent analysis examined performance differences between novices (n = 7) and experts (n = 5) under each modality.

- Under **visual-only** guidance, novices achieved a *Successful Completion* rate of 54.29%, compared to 56.00% for experts. *Missed Target* occurred in 37.14% (novices) and 36.00% (experts) of trials. *Critical Failures* occurred in 8.57% (novices) and 8.00% (experts).
- Under **multisensory guidance**, novices achieved 91.43% *Successful Completion* rate and experts 72.00%. *Missed Targets* were rare (8.57% novices; 20.00% experts).

Critical Failures were absent for novices (0.00%) but occurred in 8.00% of experts trials.

Within each modality, no statistically significant differences were found between novices and physicians (V: $\chi^2 = 0.37, p = 0.829$; MS: $\chi^2 = 5.13, p = 0.074$).

The outcome rates across expertise groups and modalities are summarized in Table 5.2 and illustrated in Figure 5.7 using a stacked bar plot.

Table 5.2: Task outcome rates for expertise group under visual-only (V) and MS modalities. All the enrolled participants (12) performed 5 tasks per modality. Experts (5) and novices (7) subgroups were also analyzed. The higher values by *Success* and the lower values by *Missed Target* and *Critical Failure*, which show better performance, are bolded.

Group	Modality	Success	Missed Target	Critical Failure
All (n=12)	V	55.00%	36.66%	8.33%
	MS	**83.33%**	**13.33%**	**3.33%**
Novices (n=7)	V	54.28%	37.14%	8.57%
	MS	**91.43%**	**8.57%**	**0.00%**
Experts (n=5)	V	56.00%	36.00%	8.00%
	MS	**72.00%**	**20.00%**	8.00%

5.3.2. Needle Tip Placement Analysis

Analyses of final needle-tip placement were restricted to successful trials only, yielding 33 samples for V and 50 samples for MS.

Table 5.3 summarizes the key descriptive statistics for both analyses, and Figure 5.8 schematically shows them. The variable n represents the number of considered samples for each modality. Given the non-normal distribution of the data, the median is reported alongside measures of dispersion, Mean Absolute Deviation (MAD) and Interquartile Range (IQR), as well as the range (Min and Max).

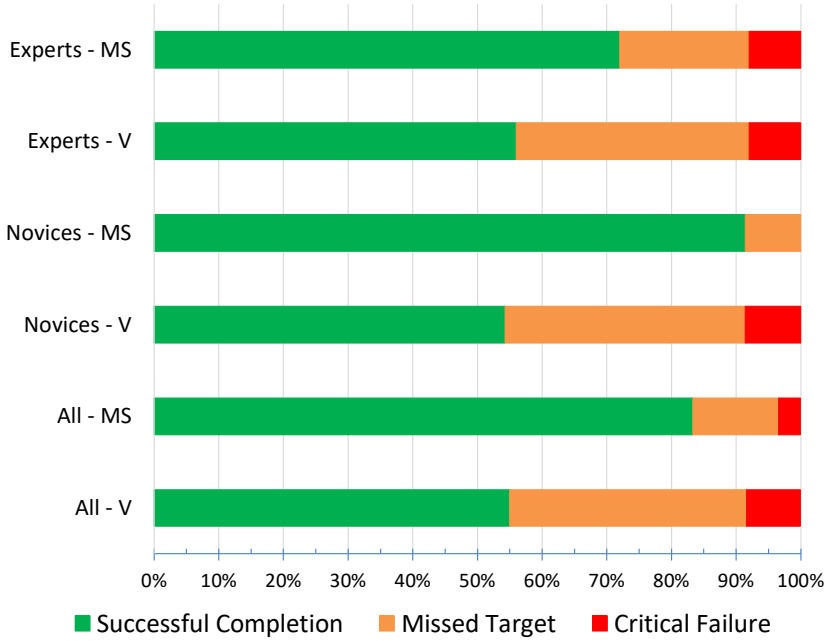


Figure 5.7: Distribution of task outcome rates (*Successful Completion*, *Missed Target*, and *Critical Failure*) under visual-only (V) and MS modalities for all participants, novices, and experts. The stacked bars highlight the higher success rates and lower error rates under MS across all groups.

Table 5.3: Summary of accuracy measurements for final needle tip placement. Values are reported for successful trials only. n is the number of samples considered for each modality.

Modality	**n**	**Median**	**MAD**	**IQR**	**Min**	**Max**
Minimum Distance to Pericardium (mm) – across 20 frames						
V	33	0.48	0.36	0.98	0.03	13.01
MS	50	0.38	0.28	0.80	0.00	2.95
Distance to Target Points (mm)						
V	33	5.64	4.89	6.34	0.53	43.64
MS	50	6.16	2.60	5.05	1.00	15.30

Distance to the Pericardium

Needle placement accuracy relative to the pericardium surface was quantified as the minimum distance between the final tip position and the pericardium mesh across the 20-frame

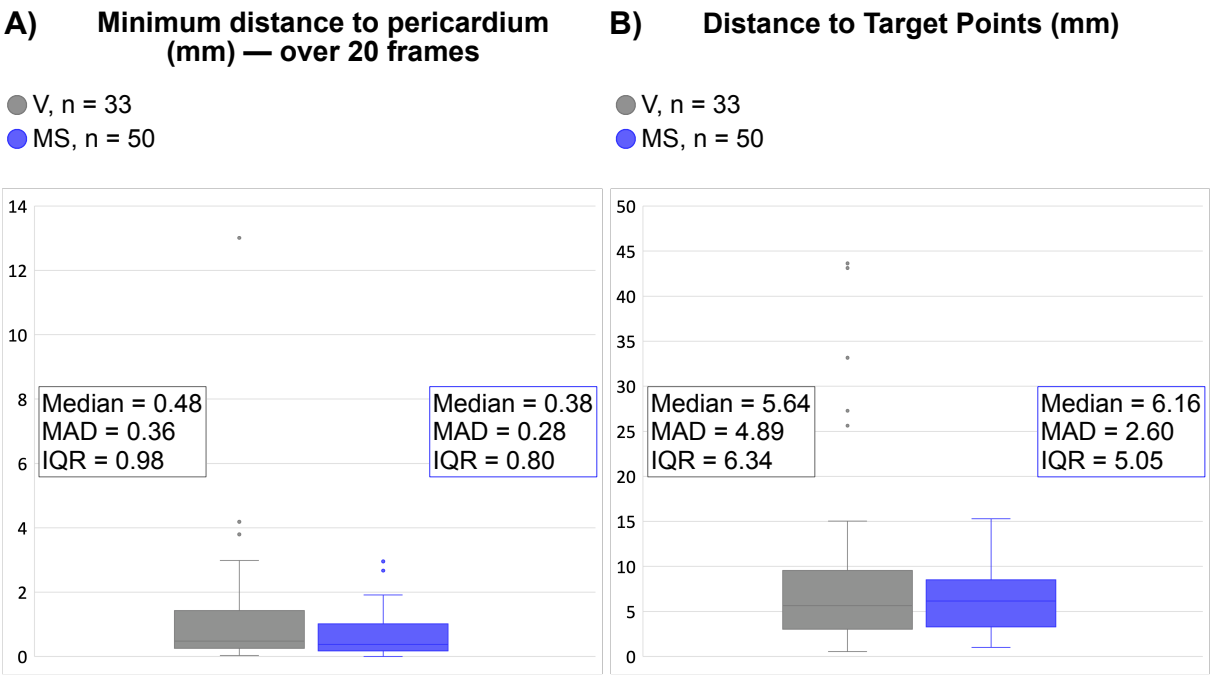


Figure 5.8: Box plot visualization of successful needle placement accuracy evaluated under two aspects : (A) stopping advancement upon reaching the pericardial membrane; and (B) accurately reaching the pre-planned target points (). The n variable represents the number of samples considered.

cardiac cycle. In this formulation, smaller values reflect higher accuracy, as they indicate that the user more promptly and reliably perceived reaching the membrane and stopped the needle, while inside the sac.

With visual-only guidance, the median distance to the pericardium was 0.48 mm (IQR: 0.98 mm, range: 0.03 mm to 13.01 mm), while under multisensory guidance it was lower at 0.38 mm (IQR: 0.80 mm, range: 0.00 mm to 2.95 mm).

A non-parametric Mann-Whitney U test was chosen because the data were not normally distributed and group sizes were unequal. A Mann-Whitney U test showed no statistically significant difference between modalities in distance to the pericardium ($U = 982, p = 0.145$), though a small effect size (Cliff's $\delta = 0.19$) suggested a marginal advantage for multisensory guidance. Fligner test indicated no statistical significance in dispersion ($p = 0.1385$).

Figure 5.9 illustrates these distributions, indicating closer proximity to the pericardium under multisensory guidance.

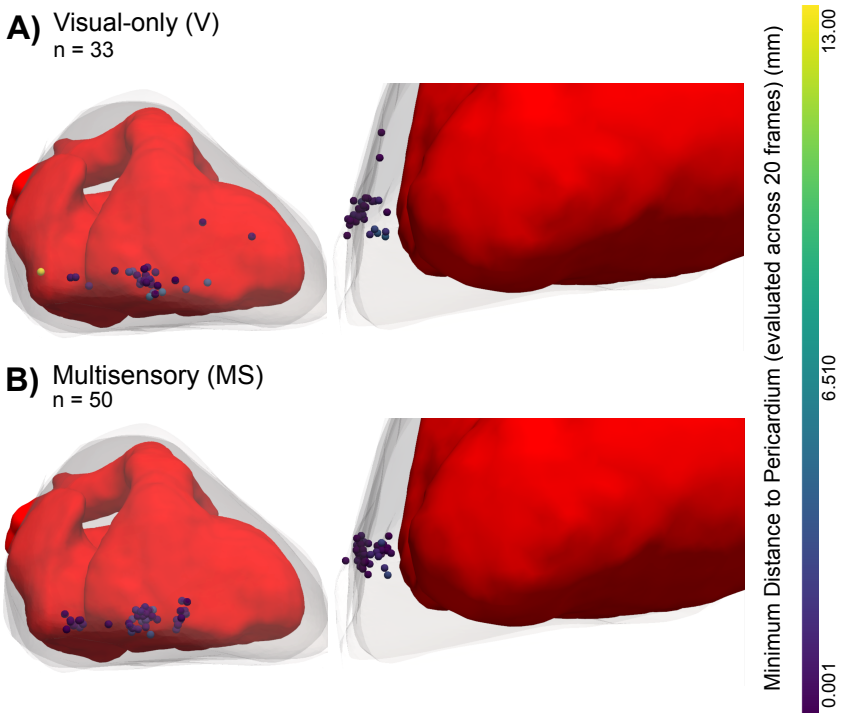


Figure 5.9: Final successful needle tip placements for both modalities relative to the pericardial membrane. Colors indicate the minimum distance to the dynamically moving pericardium (evaluated across 20 frames). The multisensory condition (bottom) exhibits greater spatial consistency, with placements more closely aligned to the membrane.

Distance to the Target Points

Further analysis was conducted on the Euclidean distance between the final needle tip position and the correspondent planned target points. These target points are defined in the End Diastolic frame which represents the most critical situation, with the myocardium at maximum volume and higher risk of myocardial injury. These points satisfied all the pre-planning safety criteria (see Section 5.2.1). Thus, lower distance to these points, ensures safer needle placement and avoids high-risk punctures. Figure 5.10 illustrates the 3D distributions of needle tip placements across modalities and targets, along with the corresponding number of samples (n). Statistical testing between the two modalities revealed no significant difference in medians (Mann-Whitney $U = 847.0000$, $p = 0.8414$) and only a negligible effect size ($\delta < 0.147$). In contrast, the Fligner test indicated a significant difference in dispersion ($p < 0.05$).

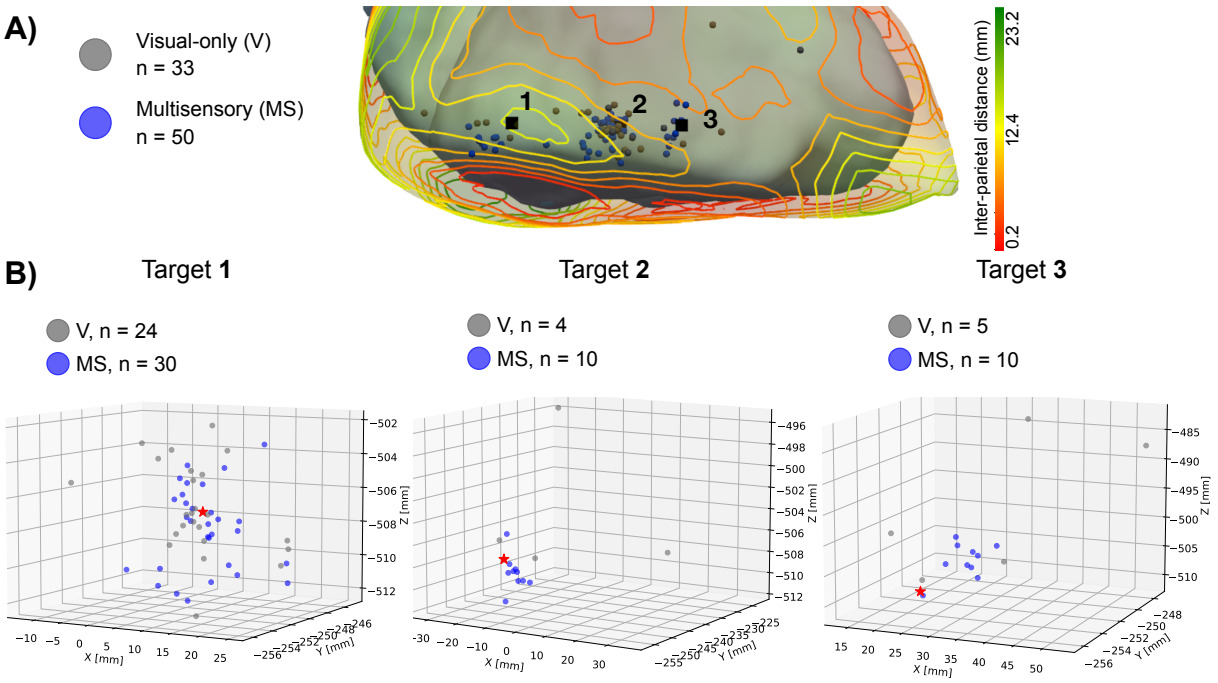


Figure 5.10: Needle tip placement relative to the planned target points 1, 2, 3 across both modalities for successful placements only. The value of considered samples for each test is reported (n). **(A)** 3D distribution of final tip positions relative to the pericardial surface (End-Diastolic phase). The clustering effect demonstrated by the MS modality, helps to reduce high-risk placements with tips avoiding critical areas of cardiac wall proximity. **(B)** Each subplot shows the spatial spread of final tips. While both modalities exhibit deviations from the planned targets, MS generally resulted in tighter clustering and reduced spread compared to V.

5.3.3. Execution Time

Execution times of the whole collected samples, recorded from task initiation to verbal completion confirmation, exhibited non-normal distributions (Shapiro-Wilk, ($p \leq 0.05$)). Puncture-related median durations were 13.82 seconds (V) and 17.57 seconds (MS). A Mann-Whitney U test showed no statistically significant difference between execution times with V modality and with MS modality ($U = 1316.0, p = 0.241$).

The correlation between successful task duration and targeting accuracy (distance to target) is shown in Figure 5.11. A modality-dependent interaction was observed (Fisher's $z = 3.12, p < 0.01$). Spearman's rank correlations revealed opposite trends:

- **Visual-only**: Positive correlation ($\rho = +0.31, p < 0.05$), indicating that longer trials were weakly associated with larger errors (lower accuracy).
- **Multisensory**: Negative corcourelation ($\rho = -0.28, p < 0.01$), suggesting that shorter trials tended to achieve higher accuracy.

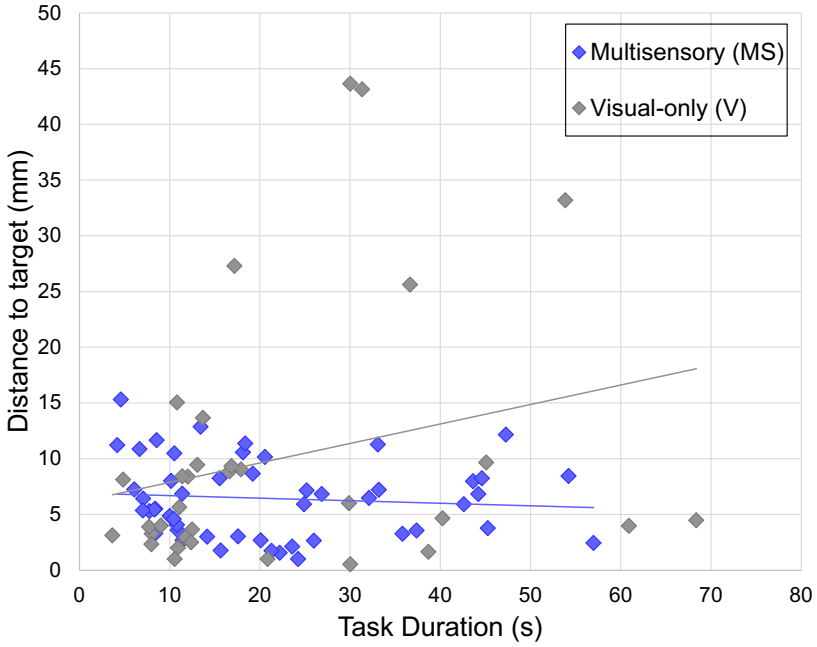


Figure 5.11: Modality-dependent time-accuracy relationships. Needle placement accuracy is measured by the distance of the needle tip position to the planned target point. The less distance represents better performance. Visual-only trials (gray) exhibit a positive correlation ($\rho = +0.31$), while multisensory trials (blue) show a negative trend ($\rho = -0.28$).

5.3.4. Workload Assessment

A paired t-test revealed a statistically significant difference in perceived difficulty between conditions ($t = -4.02, p < 0.01$), with lower difficulty ratings under MS compared to V. The analysis of the NASA-TLX questionnaire reported a relatively lower mental workload in the audiovisual condition (M = 10.00) compared to the visual-only condition (M = 12.67). While this difference was not statistically significant ($p = 0.059$), the medium-to-large effect size (Cohen's $d = 0.64$) suggests a meaningful reduction. The 95% confidence interval $[0.02, 5.32]$ further indicates that the audiovisual condition was generally perceived as less cognitively demanding.

5.3.5. Experts' Subjective Feedback

Experts strongly favored this auditory feedback approach, noting that it enabled full reliance on sound during the navigation, particularly when visual cues were compromised by registration errors or OST-HMD instability.

They highlighted that, unlike unimodal visual display, which often requires head and body movements (introducing a risk of misalignment), multisensory method offered millimeter-level precision and smooth control of the slender, highly sensitive needle.

Participants also emphasized its potential as a training tool for novices, who frequently have limited opportunities for supervised practice in PEA.

5.4. Discussion

The proposed sonification approach goes far beyond simplistic, alarm-based auditory feedback, while it introduces a novel interactive design concept based on physical modeling synthesis, specifically using a multi-layered, dynamic 2D membrane that models the physical behavior of cardiac structures. This modeling framework enables to capture and sonify continuous spatial and temporal dynamics, such as the myocardium's dynamic during diastole and systole, an aspect that to our knowledge has not been addressed in previous auditory feedback systems. Unlike conventional binary or threshold-based sonification, the system allows for expressive physically grounded audio rendering of the navigation process. The parameters chosen (e.g., radius, tension, damping) reflect real physical characteristics and contribute to the perceptual intuitiveness of the sonification.

The results support the effectiveness of the proposed methodology across all evaluated metrics, with particularly improved task completion time and usability compared to pre-

viously reported needle guidance sonification approaches based on direct parameter mapping, which have been associated with increased cognitive demand and longer learning times [74, 138].

Navigation Safety: The findings demonstrate that multisensory feedback significantly outperformed visual-only modality for a safe navigation in PEA: MS modality significantly increased success rates and reduced both missed targets and critical failures. The stacked bar visualization (Figure 5.7) highlights the marked increase in success rate and the reduction in missed targets under the multisensory condition. Notably, experts exhibited a 8% *Critical Failure* rate with both modalities, whereas novices had 8% in V and 0% with MS. This can be explained that experts did not significantly rely on the navigational cues in both modalities due to their prior expertise. Instead novices, with no experience-related biases, completed the task without any critical failures with multisensory modality. Even if the Chi-square test on the accuracy analysis conducted on expertise level subgroups doesn't indicate statical significance, the results may support the potential value of the multisensory guidance as training tool. Further studies with larger cohorts and repeated exposure to the system can substantiate these observations, and explore the interplay between expertise and multisensory integration.

Needle Placement Accuracy: The accuracy analysis of the final needle tip placement, critical for subsequent steps of the PEA procedure, yielded several insights based on the two accuracy measurements. First, regarding the distance to the pericardial membrane, both modalities were comparably accurate, with no significant difference in median distance ($p = 0.145$). However, the MS condition showed an 18% reduction in variability (IQR), indicating more consistent and predictable outcomes. In this case, users were able to guide the needle into the pericardial sac while stopping in the proximity of the outer layer, suggesting that multisensory feedback enhanced perceptual awareness of the moving target. Second, with respect to the distance from the predefined target points on the pericardial surface, MS guidance consistently reduced dispersion (lower MAD and IQR) with a tighter clustering effect around the target points, and decreased the number of outliers compared with V. Nevertheless, accuracy under this requirement remains uncertain for clinical translation, as targeting errors below 5 mm are recommended[15, 104]. It is also important to note that targeting error strongly depends on the initial needle-to-path alignment before navigation begins, since users had limited ability to adjust the trajectory once the needle was inserted into the foam chest patch. A deeper analysis that considers the entire task is therefore recommended.

Execution Time: The analysis, for successful placements only, revealed modality-specific differences in the speed-accuracy trade-off. Under V guidance, longer task dura-

tions were weakly associated with lower accuracy ($\rho = +0.31$), suggesting that additional time did not necessarily translate into more effective corrections and may instead reflect increased perceptual uncertainty. In contrast, the MS condition showed an opposite trend ($\rho = -0.28$), where longer trials tended to achieve higher accuracy. This finding supports the interpretation that multisensory feedback reduces perceptual confusion and facilitates more efficient and confident control, enabling users to succeed after a few corrective actions. The significance of the modality-dependent interaction ($p < 0.01$) confirms that MS feedback fundamentally alters the speed-accuracy relationship compared to visual-only guidance, and indicates a potentially better learnability. Additional studies will be required to further investigate this relationship between execution time and accuracy, which is expected to be influenced by longer-term user adaptation and learning effects.

The analysis of the cognitive load using NASA-TLX and perceived difficulty complements the objective accuracy results, indicating that the multisensory condition not only improved performance but also reduced perceived cognitive effort. Expert feedback further reinforces this interpretation.

These results should be interpreted in light of the primary goal, which was to demonstrate the potential of a novel multisensory guidance paradigm. Importantly, the main outcome of the proposed approach lies in its ability to improve navigation safety, representing a significant step for this application domain. Finally, further studies are needed to quantify operating room sound masking effects.

5.5. Limitations and Future Directions

While auditory feedback is the novel contribution, the scalability of the proposed XR system into real clinical settings requires careful consideration. In particular, while accurate anatomical reconstruction, robust registration, and efficient tracking are standard practice in the CAS community, as reviewed in Section 2.1, their application in this specific context may pose challenges to clinical adoption that require investigation in future work.

The clinical scalability problem can be decomposed into two requirements of different criticality thanks to the framework's modular architecture: both the sonification and visualization components are driven by the same navigation parameters yet operate independently. This decoupling in the design ensures that auditory feedback maintains high spatial accuracy regardless of the state of the visual channel.

Reliability and latency of the navigation parameters This is the primary requirement, as the navigation parameters describe the relative position of the needle tip with re-

spect to the moving cardiac anatomy and are shared across both modules. However, since they are directly mapped to the sonification output, which is the component designed for high-accuracy guidance, their accuracy and temporal consistency are of paramount importance.

This requirement depends on: accurate and real-time anatomical reconstruction of the structures of interest, temporal synchronization and spatial registration of the anatomical model with the patient's frame defined by the tracking system, and accurate and precise tool tracking.

Anatomical reconstruction: Pre-procedural ECG-gated time-resolved 4DCT enables the reconstruction of a patient's dynamic three-dimensional anatomy [110] (more details in Appendix B). By capturing CT frames over at least four cardiac cycles and assuming minimal heart rate variability, this approach provides a reliable representation of both cardiac and respiratory motion. Since the motion of critical organs depends on both cardiac and breathing contributions, and breathing is regularized by external ventilation (Section 5.1), the motion can be accurately estimated [28, 188]. The resulting dynamic model supports precise preoperative planning, facilitating the identification of safer needle landing zones based on inter-parietal distance and safe trajectories using the presented planning algorithm (Section 5.2.1).

Nonetheless, while this provides a robust basis for pre-planning, it does not account for intraoperative cardiac rhythm variability, which is not regularized like breathing. In fact, cardiac rhythm can change due to several factors, including volemia, arterial pressure, or pharmacological states during the procedure, thus affecting the reliability of intraoperative reconstruction. For this reason, a correction pipeline for such reconstructed models is presented at the end of this section.

Temporal synchronization: Temporal synchronization can be established using time-warping algorithms between the 4DCTA-based anatomical models, which are ECG-gated and therefore encode cardiac cycle information, and the ECG acquired intraoperatively [114].

Spatial registration: While advanced spatial registration methods are detailed in Section 2.1.3, none explicitly address the challenge of registering highly dynamic cardiac structures to ensure their spatial alignment with the preoperative patient anatomy and planning. For this reason, this aspect is crucial for clinical translation and requires support from an intraoperative imaging modality. A potential workflow for its integration is proposed below.

Tool tracking: Advanced tool tracking systems are widely adopted for this task and

are outlined in Section 2.1.2. While in this phantom implementation an OTS was employed, clinical translation would favor the adoption of EM tracking systems. In fact, optical tracking requires external tracking of the tool handle with tool tip position derived through a calibration pipeline, introducing potential errors. More importantly, this approach assumes the needle is completely rigid, which is a reasonable approximation for the rigid Tuohy needle and phantom used in this study, but would not be acceptable in clinical practice where even sub-millimeter deviations affect guidance and are relevant for this delicate procedure. In contrast, EM tracking enables direct tracking of the tool tip without rigidity assumptions. Moreover, sensor-equipped catheters and needles based on EM tracking technology are already widely deployed in electrophysiology [86], and several platforms have received FDA clearance, making tracking accuracy and latency well-characterized for clinical translation.

Integration of intraoperative ultrasound for motion correction: A promising future direction is the use of intraoperative 3D ultrasound (US) as a bridge between the preoperative and intraoperative domains. A 3D US volume, acquired before the procedure starts, is first registered to the 4DCTA using advanced deformable multimodal registration techniques [18, 77], establishing a common spatial reference. This reference then enables continuous intraoperative monitoring of cardiac motion through real-time registration of incoming 2D US frames [115, 218], combined with myocardium segmentation [226]. Together, these steps allow for ongoing refinement of the preoperative anatomical models and dynamic correction of the navigation parameters throughout the procedure. This approach could address both the need for intraoperative correction of the anatomically reconstructed model and the spatial registration challenge.

Latency considerations: Regarding latency, no substantial increase is anticipated under clinical conditions compared to the phantom study. If intraoperative motion correction relying on intraoperative ultrasound is adopted, the main additional contributor will be real-time image registration and segmentation in terms of model inference time, which must be carefully optimized to meet the temporal requirements of millimeter-scale navigation.

Registration to the OST-HMD The secondary requirement concerns the registration of the 4DCTA-derived anatomical models into the OST-HMD coordinate frame, ensuring that the visual overlay is correctly anchored to the patient. This challenge has been extensively explored in XR surgical navigation research and outlined in Section 2.2.3. Markerless registration based on anatomically meaningful landmarks currently represents the most suitable approach [101].

5.6. Final Remarks

In this chapter, a novel XR tool for intraprocedural guidance in PEA was presented and evaluated, where heart dynamics and the need for high precision in needle positioning pose significant difficulties. The results highlight the potential of auditory feedback to support spatial awareness and decision-making during needle navigation, particularly in dynamic anatomical environments where visual cues alone may be insufficient. While the findings of the controlled experiment including expert users indicate potential benefits of multisensory guidance, they should be interpreted as preliminary. Further evaluation in more realistic and clinically relevant environments will be important to assess robustness, usability, and practical applicability. These results provide a foundation for future research on multisensory interaction and sonification in surgical guidance.

6 |

BioSonix: A New Paradigm for Perceptually Rendering Tissue Deformations from Tool Interactions

---

This chapter is based on: Ruozzi ∗, V., Matinfar ∗, S., Schütz, L., Wiestler, B., Redaelli, A., Votta, E., & Navab, N. (2025, May). BioSonix: Can Physics-Based Sonification Perceptualize Tissue Deformations From Tool Interactions? In International Conference on Information Processing in Medical Imaging (IPMI) (pp. 19-33). Cham: Springer Nature Switzerland. This work was conducted in collaboration with the Chair for Computer Aided Medical Procedures & Augmented Reality (CAMP), TU Munich.

6.1. Perceptual Challenge of Tool-Tissue Interaction

In clinical practice, not only in the cardiovascular domain, but also in routine procedures such as needle biopsy examinations or vascular catheterizations, operators face significant perceptual challenges. In these cases, typically ultrasound-guided, the clinician must simultaneously handle the ultrasound probe and the needle, confronting substantial demands on coordination and dexterity [152].

Specifically, it is crucial to position the probe to capture the relevant anatomical plane for the puncturing action, with ultrasound image quality strictly dependent on operator manipulation [152]. At the same time, the operator must remain focused on haptic feedback, which is essential for recognizing target tissue, detecting potential damage, and avoiding complications. However, as discussed in the Introduction (Chapter 1), haptic feedback is severely attenuated, with only gross interaction forces transmitted along the needle shaft. Fine tactile information, such as tissue texture variations, compliance changes at boundaries, or high-frequency vibrations signaling perforation, is largely filtered out before reaching the operator's hand.

This clinical overview highlights the critical importance of perceiving tool-tissue interactions beyond what current visual and haptic channels can provide.

Augmentation Strategies Medical XR systems face the challenge of delivering intuitive, real-time feedback to surgeons while achieving seamless integration into the surgical workflow, even when relying on accurate deformation modeling. To overcome the limitations outlined above and those already presented in Section 2.4.1 regarding unimodal visual channel saturation, researchers have explored advanced solutions beyond visual feedback.

One approach has been to augment haptic feedback directly. Studies, such as Mostafa et al. [187], highlight benefits including faster task completion and reduced reliance on intraoperative CT scans for needle placement. However, advanced haptic systems face substantial integration challenges. Unlike passive visual or auditory feedback, bi-directional haptics raise safety concerns and complicate system design. Replicating tactile sensations (e.g., texture, temperature, pressure) is inherently challenging, while high costs, latency, and operating room constraints further limit adoption [175]

An alternative lies in leveraging the auditory modality. While the human visual system excels at spatial representation, the auditory system is remarkably adept at capturing the dynamic characteristics of objects with exceptional temporal resolution, as outlined in

Section 2.4. Operating subtly and efficiently, auditory cues serve as a cost-effective and intuitive medium for conveying intricate information.

This potential has motivated studies such as [89, 160] to explore real-time acoustic sensing techniques for capturing tool-tissue interactions. These systems employ contact microphones and acoustic emission sensors to detect high-frequency mechanical events at the tool-tissue interface. They are sensitive to interaction forces and microscale friction between the tool and surrounding tissue, capturing phenomena such as tissue layer transitions, the rupture of microscopic tissue bonds, and structural boundary crossings with exceptional temporal resolution. While promising, a critical limitation remains: the raw signals provided by these methods are often beyond the range of human hearing or lack the perceptual structure necessary for intuitive interpretation by operators. This limitation suggests sonification as a viable solution: transforming these imperceptible signals into interpretable acoustic cues through systematic mapping strategies.

However, the dynamic nature of surgery, influenced by patient-specific anatomy, procedural variability, and diverse surgical strategies, presents significant challenges for simple parametric sonification approaches. Direct mappings (e.g., force to pitch, velocity to volume) may work in controlled experimental settings but struggle to generalize across the diverse conditions encountered in clinical practice (Section 2.4.2). Augmentation systems must adapt to the nuanced demands of surgical procedures to effectively support the expertise of highly trained surgeons.

For this reason, physics-based sonification (Section 2.4.3), specifically the extension of *"From Tissue to Sound"* [135], appears promising to address this challenge by generating perceptually meaningful acoustic feedback grounded in biomechanical principles.

6.1.1. Proposed Solution

This Chapter presents an auditory-based XR tool designed to enhance nuanced decision-making in complex surgical tasks, addressing the limitations of Medical XR System in augmented soft tissue deformation. It advocates for a systematic sonification approach that moves beyond discrete tissue states, focusing on defining tool-tissue interactions to (1) develop a generalizable model adaptable to diverse scenarios and (2) provide detailed insights that enable surgeons to integrate prior knowledge, explore the surgical field, deepen anatomical understanding, and minimize noise and errors. The aim is to deliver perceptually intuitive feedback that enhances cognitive and procedural understanding while supporting decision-making.

As a proof of concept, we propose a methodology for dynamic modeling and sonification of

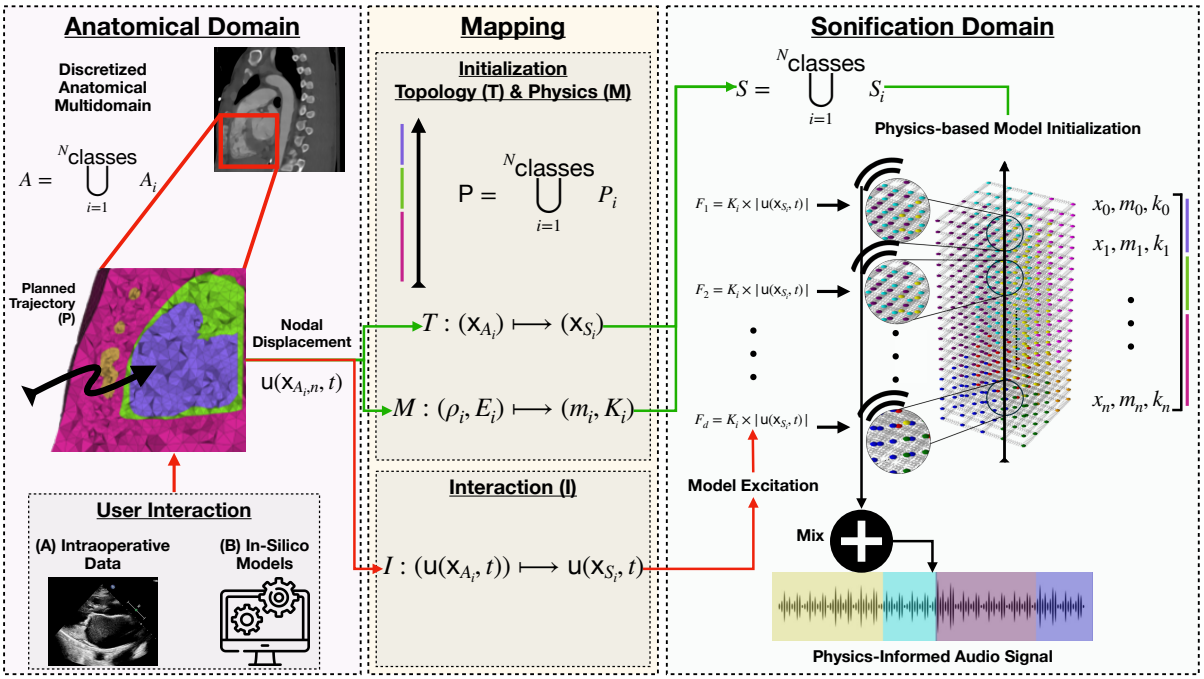


Figure 6.1: BioSonix Design Framework – A General Overview

tool-tissue interactions. The framework translates tissue displacements, reflecting tissue characteristics and interactions, into excitation forces within a physics-based sound model, enabling physics-driven auditory feedback. Biomechanical numerical simulations, specifically Finite Element Modeling (FEM), are employed to model tool-tissue interactions by probing the anatomical model across different regions and trajectories. These simulations enable the representation of diverse tool-tissue dynamics. We validate the approach through experiments analyzing the correlation between biomechanical displacement and acoustic output, optimizing the sonification model. User studies assess its technical validity, perceptual accuracy, and clinical relevance, providing guidelines for fine-tuning the model across various scenarios.

6.2. BioSonix Design Framework

The proposed framework, shown in Figure 6.1, comprises three modules: (i) the Anatomical Domain Module, which reconstructs anatomy from preoperative images while preserving tissue spatial information; (ii) the Mapping Module, which translates system physics into the Sonification Domain; and (iii) the Sonification Domain Module, which generates the final audio output. The following sections detail each module, their rationale, and interconnections.

6.2.1. Anatomical Domain Module

Surgical tool navigation strategies are typically planned using preoperative static volumetric images, such as CT scans. These images enable 3D reconstruction of the anatomical region of interest (ROI) and the identification of safe pathways to target structures while avoiding critical tissues. Advanced processing software generates multidomain spatial discretizations of the ROI, preserving tissue classes and defining an anatomical discrete domain, A.

The anatomical domain A is represented as a union of subdomains, each corresponding to a specific tissue class. Mathematically, this can be expressed as:

$$A = \bigcup_{i=1}^{N_{classes}} A_i, \quad \text{with} \quad A_i = \bigcup_{e=1}^{N_{el,i}} V_e^{(i)}. \tag{6.1}$$

where A_i is the subdomain corresponding to tissue class i, $N_{classes}$ denotes the total number of tissue classes, $V_e^{(i)}$ represents a single volumetric element e within subdomain A_i, and $N_{el,i}$ is the number of volumetric elements in A_i.

In the surgical field, the domain undergoes solicitations from sources such as physiological motions (e.g., heartbeat and respiration) and interactions between surgical tools and anatomical structures. These solicitations generate a 3D displacement field over time, which can be referred to the nodes of the domain A, and expressed as $\mathbf{u}(\mathbf{x}A_i, n, t), \quad i = 1, \ldots, N classes, \; n = 1, \ldots, N_{nodes}$, where $\mathbf{x}A_i, n$ are the spatial coordinates of the n-th node in subdomain A_i, and N$nodes$ is the total number of nodes in the anatomical domain A.

The displacement field at time t is calculated as the difference in a node's spatial coordinates between consecutive time frames

$$\mathbf{u}(\mathbf{x}_{A_i,n}, t_i) = \mathbf{x}_{A_i,n}(t_i) - \mathbf{x}_{A_i,n}(t_{i-1}), \tag{6.2}$$

or for simplicity, we will refer to the nodal displacement as $\mathbf{u}(\mathbf{x}_{A_i}, t)$, omitting the nodal index n and focusing primarily on the tissue class i.

The temporal and spatial dynamics of the anatomical model can be derived through two distinct tracks, as shown in Figure 6.1. In **Track A**, the displacement field $\mathbf{u}(\mathbf{x}A_i, t)$ is computed from intraoperatively acquired dynamic volumetric data, such as 3D ultrasound, ensuring high accuracy by capturing all motion sources. However, tool-tissue interactions

in this track are deterministic. Conversely, **Track B** uses user-driven interactions to compute $\mathbf{u}(\mathbf{x}A_i, t)$, simulating dynamic behavior induced by surgical navigation through in-silico models that integrate system physics using computational methods.

6.2.2. Mapping Module

The Mapping Module defines the methodology for utilizing the system's physics in A to: (1) initialize and (2) excite the physics-driven sound model. (1) Initialization: The sound model is initialized using two mapping functions, M and T. The function M maps the physical properties:

$$M : (\rho_i, E_i) \longmapsto (m_i, K_i). \tag{6.3}$$

Each subdomain A_i is defined by specific density values ρ_i and Young's Modulus E_i, which approximate homogenized tissue properties and linear elastic mechanical behavior. The mapping function M_i converts these values into mass m and stiffness K_i for the mass-spring model underlying the sound domain's physical representation. To represent the spatial distribution of tissue classes, a new coordinate system $\boldsymbol{p}$ is defined along the planned safe trajectory. By projecting the intersected elements $V_e^{(i)}$ of the 3D anatomical domain A onto $\boldsymbol{p}$, the domain is reduced to a 1D coordinate system. This system is divided into segments labeled by tissue classes (i), where each segment's length represents the distance traversed through that tissue. The 1D domain $\boldsymbol{P}$ is defined as:

$\boldsymbol{P} : \bigcup_{i=1}^{N_{classes}} [0, \ell_i]$,

where ℓ_i is the segment length for tissue class i. The mapping function T then maps nodes from the Anatomical Domain A to the corresponding nodes in the Sound Domain S for each tissue class i:

$$T : (\mathbf{x}_{A_i}) \longmapsto (\mathbf{x}_{S_i}), \tag{6.4}$$

2) The module for exciting forces (F) applied within the Sound Domain S, in time and in space, is computed by the function I:

$$I : (\mathbf{u}(\mathbf{x}_{A_i}, t)) \longmapsto F(\mathbf{x}_{S_i}, t) = K_i \times |\mathbf{u}(\mathbf{x}_{S_i}, t)|. \tag{6.5}$$

6.2.3. Sonification Domain Module

The Sonification Domain Module is built on a physical model that implements mass-spring interactions, functioning as a virtual instrument initialized and excited by physical data. The domain S is defined as:

$$S = \bigcup_{i=1}^{N_{classes}} S_i, \tag{6.6}$$

where S_i represents the subdomain corresponding to each tissue class i. Building on the sound physical model presented in [215], the defined coordinate system $\boldsymbol{p}$ serves as the central axis, extending along 3D orthogonal directions to form a complex network of masses interconnected by linear spring elements. The spatial distribution of tissue classes is propagated from $\boldsymbol{p}$ to neighboring regions through the mapping function T, resulting into subsets of masses and springs specific to each tissue class. Using M, each mass subset is characterized by a mass value m_i, and each spring subset is characterized by a stiffness value K_i. The virtual instrument is excited by defining a number d of *Drivers* at multiple locations around the model ($\mathbf{x}_{S_i}$). The excitation force $F(\mathbf{x}_{S_i}, t)$ applied at each driver is mapped by I. This results in multiple excited regions, generating specific wave patterns that encode the physical properties and dynamics of the anatomical domain. The signals generated from these excitation area are gathered by *Listeners* and integrated into a Mixer, which produces the final audio output with a specific wave pattern encoding the physical properties and dynamics of the anatomical domain.

6.3. Methods

To establish a robust and stable methodology for the proposed sonification framework, this study makes use of the flexibility of traditional biomechanical models, such as the Finite Element Method (FEM), to simulate user interactions within a controlled environment. FEM remains a highly versatile and reliable approach for studying tissue-tool interaction dynamics. Its strength lies in its ability to simulate a wide range of user interactions, including variations in insertion trajectories, probing digital organs with different configurations, assessing target area accessibility, and modeling distinct physical properties. This section outlines the use of FEM-based simulations to incorporate user interactions into an abstract tissue model through needle insertions. These simulations effectively captured the diverse dynamic interactions between surgical tools and tissues, facilitating the exploration of correlations between these interactions and the resulting

acoustic signals.

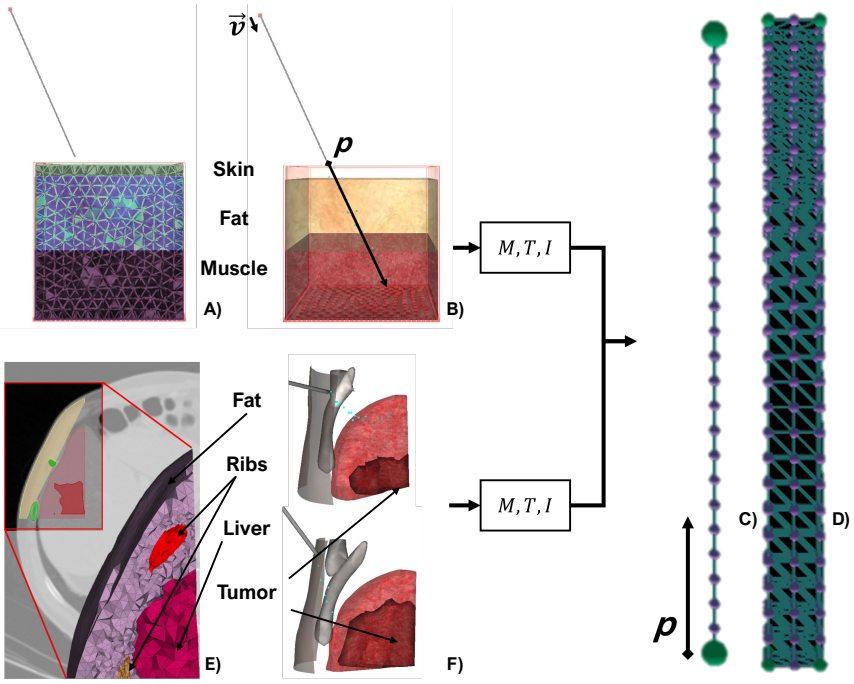


Figure 6.2: FEM-based simulations model tool-tissue dynamics by simulating insertion trajectories in a simplified setup. (A) The *Anatomical Domain* includes skin, fat, and muscle, discretized into tetrahedral elements. (B) The tool insertion is defined by an entry point, orientation, and a velocity $\vec{\boldsymbol{v}}$ applied to the handle. The $\boldsymbol{p}$-axis, aligned with the tool, spans the model, where nodal displacements $\mathbf{u}(\mathbf{x}_{A_i}, t)$ are computed along specific nodes (light blue dots) and mapped to the *Sonification Domain* via M, T, and I. (C) A 1D spring along the $\boldsymbol{p}$-axis represents the basic sound model, while (D) a more advanced model expands orthogonally, forming a 3D topology centered on the $\boldsymbol{p}$-axis.

6.3.1. Abstract Anatomical Model

The simulations were conducted using the software SOFA Framework[1] with the Cosserat Plugin[2], which is designed for the simulation of tool-tissue interactions.

The Cosserat Plugin includes an open-source implementation of needle insertion simulation, based on Cosserat Theory [13]. The needle is represented using linear elastic beam elements with six degrees of freedom. The interaction between the needle and tissue is modeled using the Complementary Constraints approach [51].

Building upon this existing FEM simulation, a non-homogeneous tissue model was in-

[1] https://www.sofa-framework.org/
[2] https://github.com/SofaDefrost/Cosserat

corporated. The anatomical model was defined as a cubic geometry with dimensions $25cm \times 25cm \times 25cm$, discretized into 13,465 tetrahedral elements. The bottom face of the cube is fixed, while a spring with a stiffness of $10^2 MPa$ is applied to the nodes on the lateral faces to prevent excessive lateral motion. The tissue model consisted of three distinct sub-domains, corresponding to three parallel layers (see Figure 6.2A), each representing a specific tissue class. In the presented preliminary study skin, fat and muscle are considered.

The constitutive model of the tissue layers were approximated using a linear elastic constitutive law, with a co-rotational formulation employed to handle large deformations. Each sub-domain was characterized by its density (ρ) and a Young's Modulus (E), with values derived from literature [190]. A uniform Poisson's ratio of 0.4 was applied across all domains.

The needle was modeled as a 17 Gauge, 26 cm-long rigid structure with a Young's modulus of 210 GPa. For the simulation, it was discretized into 10 beam elements.

Each needle insertion was simulated by specifying an entry point on the top surface of the tissue model and an initial tool orientation in space. During the simulation, the needle was pushed from its handle with a constant velocity along the predefined trajectory. The remaining degrees of freedom of the tool were unconstrained, with its configuration at each time step determined by solving the contact interaction with the tissue through the Complementary Constraint problem. Herein, the simulated trajectory was defined as shown in Figure 6.2B and a velocity of $\vec{\boldsymbol{v}} = \frac{1\,\mathrm{cm}}{s}$ is imposed.

6.3.2. Sound Model

The development of the Biosonix design framework is facilitated by the open-source miPhysics modelling libraries [3], which enable the implementation of the physics-based sound model with the characteristics described in Section 6.2.3. The $\boldsymbol{p}$-axis is defined as a straight path with the same orientation in space of the simulated trajectory traversing the whole model.

The informative nodes of the anatomical model from which the displacement is extracted ($\mathbf{u}(\mathbf{x}_{A_i}, t)$) are specific to the $\boldsymbol{p}$-axis and are mapped to the sound models by T (Figure 6.2B)). The extracted $\mathbf{u}(\mathbf{x}_{A_i}, t)$ undergoes minimal processing: the signals are first normalized within the 0.1 to 0.9 percentile range and subsequently scaled to a new range spanning from 0 to 0.1. For the preliminary benchmark, the sound model is imple-

[3] `https://github.com/mi-creative/`

mented in its simplest form: a 1D string directly projected from the $\boldsymbol{p}$-axis as shown in Figure 6.2C. The spring connections are 10 mm long. The amount of *Drivers* d and *Listeners*, and their location will be defined in Section 6.3.3.

6.3.3. Model Optimization

To provide effective perceptual feedback to the user when touching the target anatomy, it is essential to define both d and position of *Drivers* starting from the target region, where the greatest emphasis is required. In this abstract model, we consider muscle tissue as the target structure. Within this controlled simulated environment, three systematic experiments were conducted to investigate the impact of the mapping module on the resulting sound. In Figure6.3 few examples of this systematic approach are illustrated.

Nodal Characterization Given that the physics-informed subsets of masses and springs are interconnected into a unified model, it is essential to analyze the contribution of each node to the resulting sound when excited. Figure 6.3A, B, C illustrate the correlation between displacement data and spectrograms. Notably, transitions corresponding to the tool puncturing the tissue, such as from skin to fat and from fat to muscle, are highlighted by red lines in the displacement plot.

Sonic Area Definition The Sonic Area is defined as the minimal informative region above the target that contributes significant information to the resulting sound. In this abstract model, where the top surface of the muscle layer is considered the target anatomy, defining the Sonic Area is essential for determining the maximum distance between the *Drivers*. This, in turn, dictates the extent of the Sonic Area required to preserve critical information without loss. Figure 6.3D illustrates the result of selecting a narrow area above the target, while Figure 6.3E shows a wider area extending from the target to the top of the skin surface. Both configurations are valuable, but the information they convey differs and can be adapted to suit various tasks.

Nodal Contributions Impact With the Sonic Area defined, it remains to be investigated how the final sound is influenced by adding contributing nodes from different locations within the Sonic Area. This aspect is particularly important for creating smooth sound transitions as the tool approaches the target. Figure 6.3F illustrates the resulting gradient in the spectrogram, reflecting this sound transition.

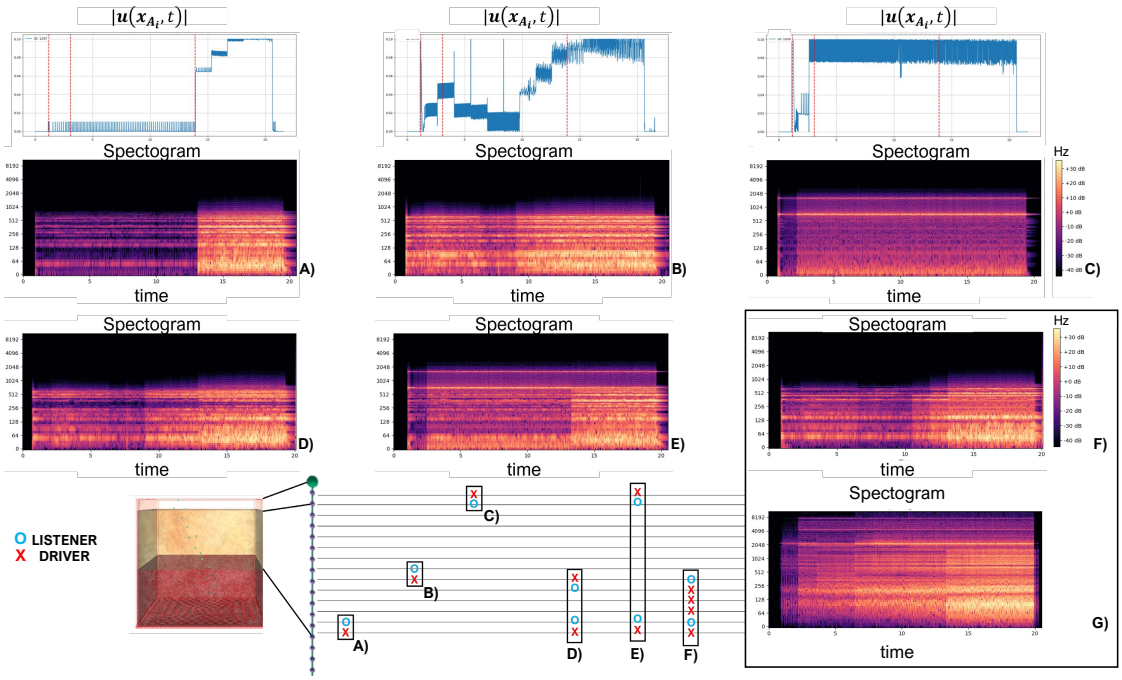


Figure 6.3: Examples from the systematic experimentation to optimize the sound model. (A, B, C) Results from the *Nodal Characterization* experiment, demonstrating the correlation between displacements in the *Anatomical Domain* and the resulting sound. (D, E) Results from the *Sonic Area Definition* experiment, illustrating the effects of varying the distance between *Drivers*. (F) A sample from the *Nodal Contribution Impact* experiment, showing the gradient achieved by activating multiple nodes.

6.3.4. Fine Tuninig

To further enhance the potential of the sound model, we introduced a more advanced design: a 3D topology formed by expanding the 1D structure along its orthogonal directions, centered on the $\boldsymbol{p}$-axis as shown in Figure6.2D. Following the in-depth and systematic investigation described in the previous section, the focus shifted to extracting the most contrasting sounds to provide users with highly perceptible auditory feedback. To achieve this, the function M, responsible for mapping physical parameters from A to S, was empirically fine-tuned. The output values returned by M, (m_i, K_i), were tested within the physics-based sound model to ensure stability and maximize perceptibility. Figure 6.3G illustrates the spectrogram generated from the same sonification configuration shown in Figure 6.3F, but with the physical parameters refined through this tuning process.

6.4. Experiments and Results

6.4.1. Toward Realistic Models

The structured method described in Section 6.3 facilitated the selection of the optimal sound model configuration. Building upon the abstract model shown in Figure 6.2B, new extensions of the model were developed by: i) incorporating additional tissue classes with varying thickness distributions within the multilayer cubic geometry; ii) implementing a new simulation setup to represent a tumor structure within a homogeneous medium; and iii) simulating diverse user interactions by varying trajectory definitions. These extensions demonstrate that the BioSonix design framework is robust when applied to this simplified abstract model across all tested scenarios. Figure 6.4 provides illustrative examples. Figure 6.4A and B highlight the stability and responsiveness of the framework when tissue classes are altered, along with the resulting changes in tissue deformability due to varying material properties. The spectrograms clearly show transition regions between tissues, with signal intensity reflecting the stiffness of the tissue probed by the tool. Additionally, a comparison between Figure 6.4B and C demonstrates the framework's ability to convey distinct dynamic tool-tissue interactions, here simulated through variations in tool trajectory, via sound signals.

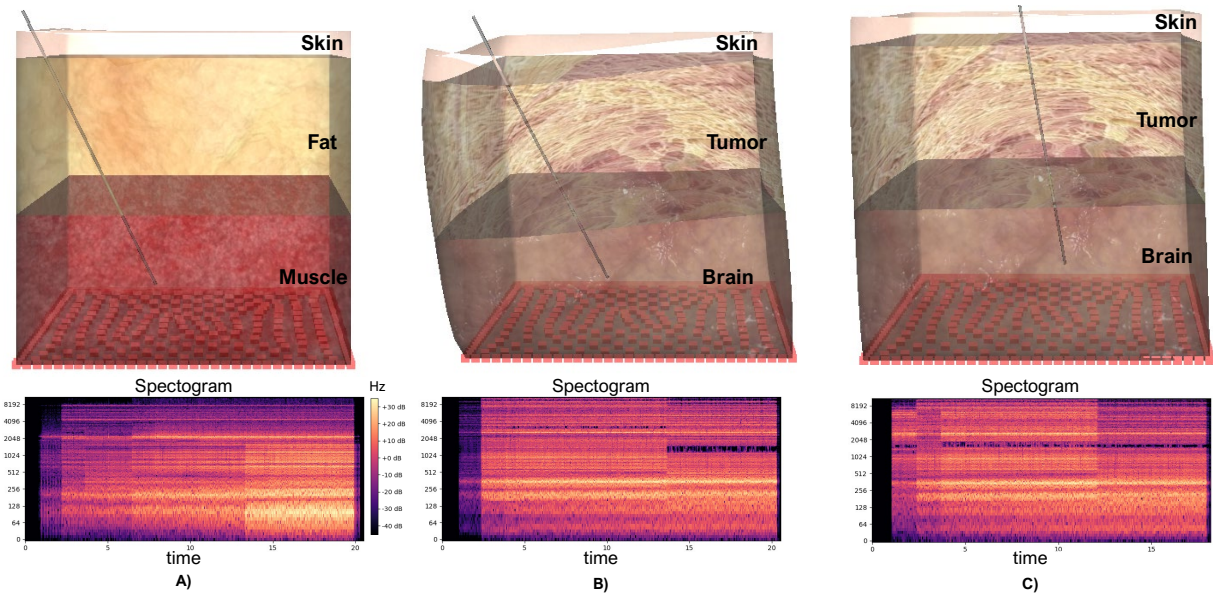


Figure 6.4: Illustrative examples of the BioSonix design framework applied to abstract models. (A, B) Demonstrate the framework's stability and responsiveness when tissue classes and their deformability vary. Spectrograms reveal transitions between tissues, with signal intensity reflecting tissue stiffness. (C) Highlights the framework's ability to differentiate dynamic tool-tissue interactions through sound signals, simulated via varied tool trajectories.

6.4.2. Real Anatomical Models

To demonstrate the applicability of the BioSonix design to real surgical tasks, the liver tumor biopsy procedure was considered. The modeling of the *Anatomical Domain A* is shown in Figure 6.2E. Starting from a public CT dataset [4], a small region of interest (ROI) encompassing subcutaneous fat, liver, and ribs was segmented. The 3D reconstruction into a multi-domain model, reflecting distinct tissue classes, was built using the ANSA v24.0.1 software (Beta CAE Systems, USA). Four different tumor configurations were created. The prepared model was imported into SOFA (Figure 6.2F) for simulating three different insertion trajectories. The sound model implemented is consistent with that described in Section 6.3.4, but the length of the spring connections was adjusted to 2 mm to match the total length of the trajectory.

6.5. User Experiments and Expert Evaluation

To assess the method's relevance from a user perspective and its applicability to realistic task scenarios and clinical settings, we conducted two user studies, as detailed below.

A video illustrating the resulting sounds and the user study design is available online[5].

6.5.1. General Perceptibility Study

The main objective of this study was to verify the perceptibility of sounds generated by the abstract models. The simulations for this study were based on the abstract model and the sound defined in Section 6.3.1 and Section 6.3.4, respectively. Both multilayer set-up and tumor structure scenario within a homogeneous tissue were implemented. In the multilayer setup, the critical layer is always assumed to be at the bottom of the model. Six tissue classes were selected based on their mechanical properties to encompass the full range of physical characteristics typical of soft tissues, as listed in Table 6.1. Additionally, bone tissue was included to introduce more variability within the physical property range. Two insertion trajectories for the needle are defined, and a constant insertion velocity of $\vec{\boldsymbol{v}} = \frac{1\text{cm}}{s}$ is imposed. In the multilayer abstract model, the thickness of each of the three tissue layers was randomized to introduce variability in the timing when the tool touches the critical layer. In the simulation with a tumorous structure within a homogeneous tissue, the tumor's location was varied. During sound generation, minimal processing was applied to each sample to manage the wide variability of tissue classes, ensuring a

[4] `https://github.com/wasserth/TotalSegmentator`
[5] See the BioSonix Design

contrasted sound across all of them. This approach allowed for the generation of 41 audio samples to conduct the study. It consisted of three stratified interactive tasks designed to be the same for all participants: (1) **Stopping:** determining when the tool enters critical tissue, (2) **Tumor Detection:** identifying whether the tool has touched the tumor, and (3) **Sorting:** sequencing the tissue classes as the tool passes through them. Before each task, participants received a brief explanation and training, including three audiovisual examples demonstrating tool-tissue interactions. Subsequently, each task consisted of 10 audio-only examples, presented in a randomized order for each participant.

Table 6.1: Mechanical properties of various tissue classes, including the elastic modulus (E) in megapascals (MPa) and density (ρ) in kilograms per cubic meter (kg/m^3).

Property	**Fat**	**Brain**	**Muscle**	**Skin**	**Tumor**	**Bone**
E [MPa]	0.02	0.036	1.6	84	100	10000
ρ **[kg/m^3]**	940	1040	1060	1100	1100	1840

General Perceptibility Results

A total of 22 participants (4 females and 20 males, aged 22 to 64 years) took part in the study. The group exhibited diverse musical backgrounds, ranging from minimal involvement to trained musicians, engaging daily. Most participants reported occasional or regular interaction with music but did not play an instrument, providing a balanced perspective for auditory task analysis. In the sorting task, 85.26% of the 190 responses were correct, while in the tumor detection task, 83% of the 200 cases were accurately identified. The statistics for the stopping task are shown in Table 6.2.

6.5.2. Expert Evaluation of the Needle Biopsy Task

The simulation set-up presented in Section 6.4.2 was implemented and tested in this user study. The primary objective was to evaluate the relevance of tasks to realistic medical scenarios in needle biopsy, specifically the **Tumor Detection** task, which required identifying whether and when the tool touched the tumor. Twelve samples were generated by combining the four tumor configurations with the three trajectories. As in previous cases, a constant insertion velocity of $\vec{v} = \frac{1\,\text{cm}}{s}$ is imposed, and only minimal processing was applied during sound generation. Before the task, participants received a brief explanation and training, which included five audiovisual examples of tool-tissue interactions. Each task then consisted of 10 audio-only examples presented in a randomized order for each participant.

Table 6.2: The statistical results of the stopping task across all cases are presented, including Exact Match (%), the proportion of exact matches as a percentage. High-Performing values are bolded based on thresholds: Mean values close to 0 ($-0.5 \leq$ Mean $\leq +0.5$), Std. Deviation below 1.0, and Exact Match (%) above 70%. Negative values indicate that stopping occurred before the target, positive values indicate stopping after the target (in seconds), and zero indicates stopping precisely at the target. The notation T# refers to the trajectory number, while the second value indicates the thickness of the critical tissue layer (in millimeters).

Statistic	Brain-Tumor		Fat-Bone		Fat-Muscle	
	T2, 80mm	**T1, 112mm**	**T1, 112mm**	**T1, 160mm**	**T1, 112mm**	**T1, 160mm**
Mean [Sec]	-1.05	0.55	1.10	-0.71	-1.38	0.75
Std. Deviation	1.19	**0.51**	1.21	**0.90**	3.76	**0.72**
Median	-2.00	1.00	1.00	0.00	0.00	1.00
Exact Match	20.00%	45.00%	15.00%	57.14%	**76.19%**	10.00%

Statistic	Muscle-Bone		Muscle-Fat		Tumor-Brain	
	T2, 112mm	**T1, 112mm**	**T2, 112mm**	**T1, 180mm**	**T1, 112mm**	**T1, 180mm**
Mean [Sec]	-0.70	0.55	-0.60	**0.05**	**-0.05**	**0.00**
Std. Deviation	**0.47**	**0.51**	**0.50**	**0.40**	**0.23**	**0.20**
Median	-1.00	1.00	-1.00	0.00	0.00	0.00
Exact Match	30.00%	45.00%	40.00%	**84.21%**	**94.74%**	**89.00%**

Expert Evaluation Results

The second user study involved two male participants: a 41-year-old neuroradiologist with limited experience in needle biopsy and a 46-year-old cardiologist who performs the task weekly. Both participants reported having a musical background, primarily as listeners, with no experience playing musical instruments. For the tumor detection task, 70% of the 20 cases were correctly identified.

6.6. Discussion

BioSonix introduced an optimized method for modeling the dynamic interactions between surgical tools and soft tissues, augmented with auditory feedback. Experimental evaluations of the models demonstrated their robustness and effectiveness in capturing the behavior of diverse tissue types, including those with highly similar physical properties, such as muscle, skin, and tumors, without relying on artificial or explicitly designed mappings; instead, the method relied purely on the inherent physical properties of the tissues. The system also proved resilient to variations in interaction angles and insertion points, offering detailed and granular insights into tool-tissue dynamics. Machine-based evaluations, including signal spectral analysis, further validated the model's stability and robustness in generating signals with reliable psychoacoustic perceptibility; by analyzing

the spectral content, fundamental psychoacoustic effects were confirmed, demonstrating the model's generalizability across different scenarios.

Following this, an initial user study was conducted to further evaluate the effectiveness and perceptibility of the sonifications in semi-interactive tasks. This foundational step, intentionally designed to be accessible to users without clinical expertise, was crucial in preparing for more advanced studies involving realistic surgical scenarios. The preliminary expert study demonstrated the method's capability to represent realistic anatomical interactions, with results achieved after minimal training. Relatively high accuracy rates were observed in both user studies, and the best performance was observed in the "Tumor-Brain" case, involving thicker tissue layers, which may be attributed to the large contrast between these tissue types and the increased thickness. The approach proved highly versatile and flexible, offering detailed insights into interactions based on variables such as tissue combinations, neighborhood effects, applied force, and tool direction, all of which influenced the resulting sounds, and users achieved these promising results even with minimal training. Together, these experiments provided valuable insights into the complexity of the model and the impact of its parameter space on the resulting sound.

6.7. Limitations and Future Directions

Despite achieving high accuracy, participants expressed a need for greater contrast between tissue classes, suggesting that they had comparatively low confidence in their decisions. The issue of improving distinguishability and clarity between tissue types, while not the primary focus of this Chapter, represents an area that could benefit from further exploration in future work; approaches such as [80, 221] could be employed to optimize the parameterization of the model, or the resulting output signal could be enhanced following [183, 184].

The same versatility that enables this richness also made tasks more challenging in simpler cases involving only three tissue types, which could have been more easily sonified using parametric approaches: this study aimed to benchmark the general concept and establish a ground truth with simple use cases, rather than to optimize for every possible scenario. Relatedly, regarding the time difference and resolution observed in the stopping task – a mean deviation of $\pm$ 1 second, equating to a 10 mm tool pathway distance within a $25cm \times 25cm \times 25cm$ tissue sample – this study aimed to demonstrate the proof of concept; a detailed discussion of accuracy and resolution is therefore not entirely relevant at this stage, and more detailed studies are needed to draw accurate conclusions regarding these variables. As sonification represents a relatively novel modality for interactive tasks of this

nature, its full potential will require further comprehensive studies to be fully realized and understood. Once the application context and safety requirements are precisely defined, the region of interest (ROI) and sonic area can be optimized to meet those specific needs.

Future research should investigate more diverse and complex interaction scenarios, utilizing the method's potential to deliver richer and more dynamic feedback, especially in intraoperative interventional tasks. A particularly important direction is the transition of this FEM-based benchmark from the controlled, offline setting presented in this Chapter towards intraoperative imaging modalities, where model inputs would be driven by real-time sensing rather than precomputed simulation. In this respect, the transferability of the proposed methodology to imaging modalities such as ultrasound and intraoperative optical coherence tomography (iOCT) represents a promising avenue, as both can provide the kind of local, tissue-resolved information the model relies on to characterize tool-tissue interaction. Beyond imaging, complementary sensing strategies such as vibroacoustic (VA) sensing [212] could offer a further valuable direction: by directly capturing the mechanical vibrations generated at tool-tissue interfaces, VA sensing has the potential to provide ground-truth data on the actual interaction dynamics, paving the way for learning-based methodologies to define the parameterization and mapping of the sonification design, complementing or extending the purely physics-driven strategy adopted in this Chapter. Future work could further explore the integration of multimodal data, such as CT and ultrasound, to harness the strengths of both modalities; combining these inputs would make it possible to generate synchronized audiovisual outputs, enhancing the overall user experience and utility in surgical environments.

6.8. Final Remarks

Overall, the results presented in this Chapter highlight the method's potential for achieving precise and reliable tool augmentation, establishing BioSonix as a versatile framework for sonifying tool-tissue interactions from first physical principles and dynamic features rather than artificial mappings. Taken together, these findings provide a solid benchmark on simple use cases and lay the foundation for extending the framework toward the richer, intraoperative scenarios discussed above.

7 | Conclusions and Future Directions

7.1. Thesis Contribution

The Introduction (Chapter 1) established cardiovascular PIs as a fundamental *spatiotemporal* problem for surgical navigation: providing reliable spatial information requires updating strategies capable of representing tissue deformation, while the temporal dynamics of cardiac motion demand that such updates are conveyed to the surgeon in real-time to preserve guidance validity. This dual requirement defined the central challenge addressed throughout this thesis: developing physics-based modeling strategies and designing multisensory displays that enable XR-based surgical navigation systems to overcome the observational constraints inherent to fluoroscopy and ultrasound guidance while meeting the perceptual demands imposed by continuous cardiac motion, limited haptic feedback and already-saturated visual attention.

The fundamental contribution introduced in Chapter 3 was the explicit modeling of *surgical interaction* itself, not merely anatomical geometry or its deformation. This interaction-centered approach recognizes that tool-tissue interaction encodes properties that transcend anatomical representation alone, with potential impact for sense augmentation.

Within the Medical XR Framework, modeling these interactions produces two distinct types of information: *anatomical updates* that maintain spatial correspondence when direct observation is incomplete, and *interaction features* that enable perceptual delivery through complementary visual and auditory modalities.

This dual-output architecture addresses both the spatial challenge of predicting what cannot be directly observed and the perceptual challenge of delivering high-temporal-resolution guidance through channels suited to dynamic information.

The three presented chapters exemplify a research trajectory along the path of surgical interaction modeling, following an *amplificatio*: from anatomical updates leveraging

sensor-derived interaction data to substitute missing imaging information, to explicit investigation of the interaction's perceptual representation, culminating in a generalizable framework that formalizes the coupling between biomechanical simulation and auditory synthesis.

Chapter 4 addresses catheter navigation in transcatheter procedures, where vessel deformations induced by catheter contact remain invisible in standard fluoroscopy. The chapter introduces a framework integrating catheter sensing with interactive FEM simulation and XR visualization. Real-time catheter shape reconstruction provides sensor-derived boundary conditions that drive vessel deformation simulation. The interaction is modeled as the direct effect of the known catheter shape on the vessel walls, demonstrating feasibility of a framework where physics-based models can maintain spatial correspondence by substituting missing imaging data with sensor observations and biomechanical prediction. The digital representation of the interaction is presented to the user through visual modality via an OST-HMD. Performance and accuracy were evaluated in an experimental phantom setup under controlled catheter insertion conditions.

Chapter 5 addresses needle access toward the beating heart in percutaneous cardiac access, where continuous cardiac motion poses significant challenges for unimodal visual XR systems and requires high temporal resolution guidance. The chapter presents a multisensory XR navigation system combining temporal cardiac modeling from preoperative 4DCT with physics-driven sonification. The surgical interaction is modeled through needle proximity to cardiac layers, where each anatomical layer is physically modeled as a resonant membrane. Needle proximity data excites these membranes, synthesizing auditory feedback that encodes distance to moving cardiac structures through membrane-based acoustic synthesis, while visual display presents dynamic cardiac anatomy animating through the cardiac cycle. The digital representation of the interaction is presented through complementary visual and acoustic display modalities. The surgical navigation system was evaluated in a phantom study with 12 cardiologists, assessing both quantitative performance metrics (time, accuracy) and subjective user experience.

Chapter 6 presents the BioSonix framework, formalizing a three-module pipeline that directly couples biomechanical tool-tissue interaction modeling with acoustic synthesis. Tissue displacements computed from interactive simulation drive excitation forces in physics-based sound models, generating acoustic output that naturally reflects tissue heterogeneity and interaction dynamics. This formalization moves beyond procedure-specific implementations to establish design principles transferable across surgical domains for presenting physics-based interaction models through auditory display. The new paradigm was validated through a comprehensive user study.

7.2. General Conclusions

While grounded in specific cardiovascular use cases, transcatheter endovascular navigation and needle-based cardiac access, the methods and technologies presented in this thesis are designed to be transferable across the broader landscape of PIs. The interaction-centered modeling framework and its dual-output architecture (anatomical updates and interaction features) address challenges that extend beyond cardiovascular procedures: the observational constraints inherent to minimally invasive techniques, where instruments are visible only through limited imaging modalities while tissue deformations and mechanical interactions remain hidden, and the perceptual demands of dynamic surgical environments, where critical guidance information must be delivered under already-saturated visual attention and high-stress conditions.

The integration of auditory display alongside visual augmentation addresses a fundamental limitation of unimodal XR systems. The visual system excels at spatial organization but exhibits well-documented constraints in perceiving textural qualities and tracking multiple dynamic processes simultaneously [47]. In surgical environments, cluttered displays and visual distractions further complicate navigation under time pressure [46, 127]. Sonification exploits the temporal sensitivity and omnidirectionality of human audition to convey high-frequency changes, cardiac motion, proximity variations, contact dynamics, independent of gaze direction. When grounded in physics-based interaction modeling, auditory feedback naturally encodes mechanical properties, tissue compliance and frictional resistance, that are difficult to represent visually, particularly when those properties vary continuously over time. The complementarity demonstrated in the presented work suggests that multimodal XR systems leverage the strengths of each sensory channel: vision for spatial understanding and anatomy, audition for temporal dynamics and interaction properties.

In conclusion, the approaches proposed in this thesis advance the state of the art in Medical XR and sonification presented in Section 2.2 and Section 2.4, respectively. By establishing validated methodologies for physics-based interaction representation and anatomical updating in the absence of complete imaging, and by demonstrating that multisensory interfaces can effectively complement purely visual paradigms, this work contributes both to the technical advancement of the CAS research community and to a broader understanding of how physics-grounded, multisensory interfaces can support safer, more precise, and more intuitive surgical procedures.

At the same time, several challenges remain open regarding clinical translatability and applicability to the intraoperative scenario. These aspects define important directions

for future research and represent critical steps toward the adoption of multisensory XR guidance systems in routine clinical practice.

7.2.1. Translating Research into Clinical Practice

Translating the developed approaches from research to clinical practice requires advancing the presented methods beyond the controlled conditions of phantom studies and experimental setups. While procedure-specific technical challenges have been addressed within their respective chapters, all three frameworks share a common prerequisite for clinical adoption: integration with intraoperative imaging, particularly ultrasound. This integration serves three critical purposes for clinical translation.

First, **scalability into clinical workflows demands accurate registration**. The rigid marker-based registration strategies employed in controlled experimental setups are insufficient for clinical deployment, where patient positioning and anatomical geometry change between preoperative imaging and the procedure. Deformable CT-ultrasound registration [68, 77] would enable construction of updated anatomical models from intraprocedural ultrasound acquisitions, accommodating both tissue deformation and patient motion. This requirement applies equally to vascular and cardiac procedures: while preoperative imaging segmentation provides anatomical models, clinical deployment requires intraprocedural imaging to verify spatial alignment, and account for respiratory- and cardiac-induced motion not captured in standard 3D acquisitions. Image-based registration approaches (Section 2.1.3) therefore remain crucial for the applicability of physics-based methods in clinical environments, and their integration with interaction modeling frameworks requires systematic investigation in future work.

Second, **patient-specific calibration of physical models** requires intraoperative imaging. Physics-based models employed in the presented frameworks rely on assumed mechanical properties, as pragmatic engineering choices when patient-specific tissue property maps are not clinically available from standard preoperative imaging. The sensitivity analyses demonstrated that such assumptions remain acceptable when the primary output is anatomical configuration in presence of kinematic boundary conditions. However, patient-specific calibration becomes critical when extending toward dynamic interaction features such as contact forces, frictional resistance, or tissue compliance variations. Intraprocedural ultrasound could enable this calibration through non-invasive measurements of vessel compliance, tissue distensibility, and acoustic impedance [31], improving both biomechanical force prediction accuracy and the fidelity of auditory feedback encoding mechanical properties. This integration positions ultrasound not merely as a registration modality

but as a source of boundary condition estimation and material parameter identification for physics-based models, enabling personalized model calibration within clinical workflows when dynamic interaction features are required.

Third, **validation and operator trust** require visual confirmation through familiar imaging modalities. Interventionalists are trained on fluoroscopy and ultrasound; they expect to "see" anatomical structures and device positions through those standard images. The challenge is to *augment* these modalities, not replace them. All proposed approaches are designed to complement, not substitute, current image-based guidance. Physics-based predictions of tissue deformation provide information invisible to fluoroscopy alone but should be validated against intraprocedural ultrasound when available. Similarly, auditory feedback extends guidance into the temporal domain but operates alongside visual display, not in isolation. This complementary architecture is feasible precisely because ultrasound is non-harmful, portable, and increasingly accessible in interventional suites. Establishing trust in physics-based predictions requires demonstrating consistency with observable imaging data where such data exist, while physics-based models provide value by predicting what imaging cannot directly reveal.

7.3. Future Directions

As remarked in Section 7.2.1, the integration of intraoperative ultrasound plays a central role in advancing toward clinical applicability under several aspects. For this reason, future work should investigate tight coupling between the presented methodologies and intraoperative imaging, where imaging data continuously informs and refines physics-based models.

Beyond these functional requirements, a deeper research opportunity emerges in developing methodologies that systematically combine image-based observations with physics-based predictions within a unified computational framework. The BioSonix framework represents the most mature foundation for this integration. As outlined in the general overview (Section 6.2), the anatomical domain could be modeled from intraoperative imaging data (Figure 6.1A). The presented study was based on a controlled abstract model with displacements obtained from FEM simulations, a method which would not be suitable for intraoperative scenarios. The applicability of BioSonix to the intraoperative field would come through Track A: anatomical domain modeled from intraoperative data. This means incorporating image-derived displacement fields.

Computer vision techniques, particularly optical flow algorithms, enable tracking of tissue motion and deformation from ultrasound sequences. Traditional displacement estimation

methods [85, 92, 178] require careful per-application tuning and are sensitive to decorrelation and noise typical of ultrasound data, with frame-to-frame tracking suffering from cumulative drift during fast or large tissue movements [129]. Deep learning approaches have improved accuracy by learning displacement estimation directly from data [24, 25, 49, 232]. However, physics-based models offer a complementary pathway: rather than treating displacement estimation as purely data-driven, biomechanical constraints can regularize the inverse problem, ensuring displacement fields satisfy conservation laws and constitutive relationships [207]. This positions physics-based models not as predictors competing with imaging, but as priors that constrain the solution space to mechanically plausible deformations.

Another promising direction exploits the inherently local nature of ultrasound motion. In B-mode ultrasound, tissue motion manifests as local shifts in speckle patterns, where displacement at each pixel depends primarily on its immediate neighborhood. This suggests that methods built on local update rules with iterative refinement could capture displacement fields effectively, without relying on costly global feature extraction of encoder-decoder architectures. Recent advances in Neural Cellular Automata (NCA) [149] demonstrate promise in this direction. NCA architectures learn local update rules on grid-structured data, producing complex global patterns through iterative application, with applications in medical imaging including segmentation, classification, temporal modeling, and physical process simulation [45, 96, 97, 141, 158]. Key properties include parameter efficiency, adjustable computation at inference time, and inherent locality. To our knowledge, NCAs have not been applied to ultrasound displacement or optical flow estimation, yet their properties align well with the local nature of ultrasound motion. Investigating NCAs as a bridge between image-derived observations and physics-based regularization could yield hybrid frameworks that leverage the strengths of both data-driven and physics-driven approaches.

Beyond displacement tracking, a more fundamental research direction involves treating ultrasound imaging itself as a physical phenomenon governed by wave propagation principles. Rather than processing B-mode images as the primary data source, accessing raw radiofrequency ultrasound signals and modeling acoustic wave interaction with tissue interfaces could enable direct estimation of mechanical properties, tissue boundaries, and contact events. Establishing methodologies that exploit radiofrequency data within physics-based frameworks, potentially through differentiable ultrasound simulation coupled with biomechanical models, represents an unexplored direction that could fundamentally enhance the integration between sensing and prediction in image-guided interventions.

Collectively, these directions, including the integration of intraoperative imaging through advanced displacement tracking, physics-based regularization, and the exploitation of raw ultrasound signals, represent natural extensions of the interaction-centered modeling paradigm established in this thesis.

Bibliography

[1] N. Abi-Jaoudeh, J. Kruecker, S. Kadoury, H. Kobeiter, A. M. Venkatesan, E. Levy, and B. J. Wood. Multimodality image fusion–guided procedures: technique, accuracy, and applications. *Cardiovascular and interventional radiology*, 35(5):986–998, 2012.

[2] N. Abi-Jaoudeh, H. Kobeiter, S. Xu, and B. J. Wood. Image fusion during vascular and nonvascular image-guided procedures. *Techniques in vascular and interventional radiology*, 16(3):168–176, 2013.

[3] S. Afzal, V. Veulemans, J. Balzer, T. Rassaf, K. Hellhammer, A. Polzin, M. Kelm, and T. Zeus. Safety and efficacy of transseptal puncture guided by real-time fusion of echocardiography and fluoroscopy. *Netherlands Heart Journal*, 25(2):131–136, 2017.

[4] D. Aigner, N. Wang, D. Kielmayer, J. Steiner, J. Hochpöchler, C. Heinzl, D. Roth, F. Maurer, and C. Anthes. Cardiac visualisation along the rv-continuum-a high-fidelity pilot study. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 675–680. IEEE, 2023.

[5] Z. Al Adas, D. Uceda, A. Mazur, K. Zehner, C. J. Agrusa, G. Wang, and D. B. Schneider. Safety and learning curve of percutaneous axillary artery access for complex endovascular aortic procedures. *Journal of Vascular Surgery*, 79(3):487–496, 2024.

[6] O. Al-Ahmad, M. Ourak, J. Van Roosbroeck, J. Vlekken, and E. Vander Poorten. Improved fbg-based shape sensing methods for vascular catheterization treatment. *IEEE Robotics and Automation Letters*, 5(3):4687–4694, 2020.

[7] O. Al-Ahmad, M. Ourak, J. Vlekken, and E. Vander Poorten. Force control with a novel robotic catheterization system based on braided sleeve grippers. *IEEE Transactions on Medical Robotics and Bionics*, 5(3):602–613, 2023.

[8] A. Alharbi, A. Abanomy, M. Almeshari, H. Alahmad, A. M. Alshahrani, M. Alarifi, M. Almanaa, and K. Alenazi. Radiation protection in cardiac catheterization: In-

sights from a survey of interventional cardiologists. *Journal of Radiation Research and Applied Sciences*, 18(1):101226, 2025.

[9] P. Ambrosini, I. Smal, D. Ruijters, W. J. Niessen, A. Moelker, and T. van Walsum. 3d catheter tip tracking in 2d x-ray image sequences using a hidden markov model and 3d rotational angiography. In *Workshop on Augmented Environments for Computer-Assisted Interventions*, pages 38–49. Springer, 2015.

[10] P. Ambrosini, D. Ruijters, W. J. Niessen, A. Moelker, and T. van Walsum. Fully automatic and real-time catheter segmentation in x-ray fluoroscopy. In *International conference on medical image computing and computer-assisted intervention*, pages 577–585. Springer, 2017.

[11] R. D. Anderson, G. Lee, M. Prabhu, C. J. Patrick, I. Trivic, T. Campbell, C. K. Chow, J. M. Kalman, and S. Kumar. Ten-year trends in catheter ablation for ventricular tachycardia vs other interventional procedures in australia. *Journal of Cardiovascular Electrophysiology*, 30(11):2353–2361, 2019.

[12] C. M. Andrews, A. B. Henry, I. M. Soriano, M. K. Southworth, and J. R. Silva. Registration techniques for clinical applications of three-dimensional augmented reality devices. *IEEE journal of translational engineering in health and medicine*, 9: 1–14, 2020.

[13] S. S. Antman. The special cosserat theory of rods. In *Nonlinear problems of elasticity*, pages 259–324. Springer, 1995.

[14] A. Arya, L. Di Biase, V. Bazán, A. Berruezo, A. d'Avila, P. Della Bella, A. Enriquez, M. Hocini, J. Kautzner, H.-N. Pak, et al. Epicardial ventricular arrhythmia ablation: a clinical consensus statement of the european heart rhythm association of the european society of cardiology and the heart rhythm society, the asian pacific heart rhythm society, the latin american heart rhythm society, and the canadian heart rhythm society. *Europace*, 27(4):euaf055, 2025.

[15] A. Aryana, R. Tung, and A. d'Avila. Percutaneous epicardial approach to catheter ablation of cardiac arrhythmias. *Clinical Electrophysiology*, 6(1):1–20, 2020.

[16] H. D. Ashkezari, A. Mirbagheri, S. Behzadipour, and F. Farahmand. A mass-spring-damper model for real time simulation of the frictional grasping interactions between surgical tools and large organs. *Scientia Iranica. Transaction B, Mechanical Engineering*, 22(5):1833, 2015.

[17] T. Ates, I. H. Sukur, F. Ok, M. G. Arikan, and N. Akdogan. Global research

trends in minimally invasive treatments for kidney stones: A bibliometric analysis (2015–2024). *Urolithiasis*, 53(1):116, 2025.

[18] M. F. Azampour, M. Tirindelli, J. Lameski, M. Gafencu, E. Tagliabue, E. Fatemizadeh, I. Hacihaliloglu, and N. Navab. Anatomy-aware computed tomography-to-ultrasound spine registration. *Medical Physics*, 51(3):2044–2056, 2024.

[19] M. Bergholz, M. Ferle, and B. Weber. The benefits of haptic feedback in robot assisted surgery and their moderators: a meta-analysis. sci rep 13 (1): 19215, 2023.

[20] D. Black, C. Hansen, A. Nabavi, R. Kikinis, and H. Hahn. A survey of auditory display in image-guided interventions. *International journal of computer assisted radiology and surgery*, 12(10):1665–1676, 2017.

[21] S. Bobade and S. Asutkar. Current trends and future directions in surgery: A brief scoping review. *Multidisciplinary Reviews*, 8(1):2025028–2025028, 2025.

[22] J. Bogaert and M. Francone. Pericardial disease: value of ct and mr imaging. *Radiology*, 267(2):340–356, 2013.

[23] M. B. Boubaker, M. Haboussi, J.-F. Ganghoffer, and P. Aletti. Finite element simulation of interactions between pelvic organs: predictive model of the prostate motion in the context of radiotherapy. *Journal of biomechanics*, 42(12):1862–1868, 2009.

[24] S. Boukhtache, K. Abdelouahab, F. Berry, B. Blaysat, M. Grédiac, and F. Sur. When deep learning meets digital image correlation. *Optics and Lasers in Engineering*, 136:106308, 2021. doi: 10.1016/j.optlaseng.2020.106308.

[25] S. Boukhtache et al. A lightweight convolutional neural network as an alternative to DIC to measure in-plane displacement fields. *Optics and Lasers in Engineering*, 161:107367, 2023.

[26] N. G. Boyle and K. Shivkumar. Epicardial interventions in electrophysiology. *Circulation*, 126(14):1752–1769, 2012.

[27] A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[28] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the aapm radiation therapy committee task group no. 132. *Medical physics*, 44(7):e43–e76, 2017.

[29] A. Brost, A. Wimmer, R. Liao, J. Hornegger, and N. Strobel. Catheter tracking: Filter-based vs. learning-based. In *Joint Pattern Recognition Symposium*, pages 293–302. Springer, 2010.

[30] S. Bruns, J. M. Wolterink, T. P. van den Boogert, J. H. Runge, B. J. Bouma, J. P. Henriques, J. Baan, M. A. Viergever, R. N. Planken, and I. Išgum. Deep learning-based whole-heart segmentation in 4d contrast-enhanced cardiac ct. *Computers in biology and medicine*, 142:105191, 2022.

[31] S. Cappelletti, A. Caimi, A. Caldiroli, I. Baroni, E. Votta, S. A. Riboldi, M. M. Marrocco-Trischitta, A. Redaelli, and F. Sturla. Non-invasive estimation of vascular compliance and distensibility in the arm vessels: a novel ultrasound-based protocol. *Quantitative Imaging in Medicine and Surgery*, 12(7):3515, 2022.

[32] F. Chadebecq, L. B. Lovat, and D. Stoyanov. Artificial intelligence and automation in endoscopy and surgery. *Nature Reviews Gastroenterology & Hepatology*, 20(3): 171–182, 2023.

[33] W.-Y. Chan and P.-A. Heng. Visualization of needle access pathway and a five-dof evaluation. *IEEE journal of biomedical and health informatics*, 18(2):643–653, 2013.

[34] P.-L. Chang, A. Rolls, H. De Praetere, E. Vander Poorten, C. V. Riga, C. D. Bicknell, and D. Stoyanov. Robust catheter and guidewire tracking using b-spline tube model and pixel-wise posteriors. *IEEE Robotics and Automation Letters*, 1(1): 303–308, 2016.

[35] E. D. Chase, T. Gerstenberg, and S. Follmer. Realism of visual, auditory, and haptic cues in phenomenal causality. In *2023 IEEE World Haptics Conference (WHC)*, pages 306–312. IEEE, 2023.

[36] F. Chen, X. Cui, J. Liu, B. Han, X. Zhang, D. Zhang, and H. Liao. Tissue structure updating for in situ augmented reality navigation using calibrated ultrasound and two-level surface warping. *IEEE Transactions on Biomedical Engineering*, 67(11): 3211–3222, 2020.

[37] J. Chen, Y. Liu, S. Wei, Z. Bian, S. Subramanian, A. Carass, J. L. Prince, and Y. Du. A survey on deep learning in medical image registration: New technologies, uncertainty, evaluation metrics, and beyond. *Medical Image Analysis*, 100:103385, 2025.

[38] K. Cleary and T. M. Peters. Image-guided interventions: technology review and clinical applications. *Annual review of biomedical engineering*, 12(1):119–142, 2010.

[39] S. Condino, R. Piazza, M. Carbone, J. Bath, N. Troisi, M. Ferrari, and R. Berchiolli. Bioengineering, augmented reality, and robotic surgery in vascular surgery: A literature review. *Frontiers in Surgery*, 9:966118, 2022.

[40] P. R. Cook. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal*, 21(3):38–49, 1997.

[41] H. Courtecuisse, H. Jung, J. Allard, C. Duriez, D. Y. Lee, and S. Cotin. Gpu-based real-time soft tissue deformation with cutting and haptic feedback. *Progress in biophysics and molecular biology*, 103(2-3):159–168, 2010.

[42] E. M. Cronin, F. M. Bogun, P. Maury, P. Peichl, M. Chen, N. Namboodiri, L. Aguinaga, L. R. Leite, S. M. Al-Khatib, E. Anter, et al. 2019 hrs/ehra/aphrs/lahrs expert consensus statement on catheter ablation of ventricular arrhythmias. *EP Europace*, 21(8):1143–1144, 2019.

[43] A. d'Avila, A. Aryana, V. Y. Reddy, and F. E. Marchlinski. *Percutaneous Epicardial Interventions:: A Guide for Cardiac Electrophysiologists*. Cardiotext Publishing, 2020.

[44] B. Desjardins and E. A. Kazerooni. Ecg-gated cardiac ct. *American Journal of Roentgenology*, 182(4):993–1010, 2004.

[45] M. Deutges, A. Sadafi, N. Navab, and C. Marr. Neural cellular automata for lightweight, robust and explainable classification of white blood cell images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 693–702. Springer, 2024.

[46] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy*, 27(2):454–461, 2013.

[47] B. J. Dixon, M. J. Daly, H. H. Chan, A. Vescan, I. J. Witterick, and J. C. Irish. Inattentional blindness increased with augmented reality surgical navigation. *American journal of rhinology & allergy*, 28(5):433–437, 2014.

[48] R. Docea, J. Xu, W. Ling, A. C. Jenke, F. R. Kolbinger, M. Distler, C. Riediger, J. Weitz, S. Speidel, and M. Pfeiffer. Seesaw: Learning soft tissue deformation from laparoscopy videos with gnns. *IEEE Transactions on Biomedical Engineering*, 71 (12):3432–3445, 2024.

[49] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox. FlowNet: Learning optical flow with convo-

lutional networks. In *International Conference on Computer Vision (ICCV)*, pages 2758–2766, 2015.

[50] C. Duriez. *Real-time haptic simulation of medical procedures involving deformations and device-tissue interactions*. PhD thesis, Université des Sciences et Technologie de Lille-Lille I, 2013.

[51] C. Duriez, C. Guébert, M. Marchal, S. Cotin, and L. Grisoni. *Interactive Simulation of Flexible Needle Insertions Based on Constraint Models*, pages 291–299. 2009. doi: 10.1007/978-3-642-04271-3_36.

[52] M. O. Ernst and M. Di Luca. Multisensory perception: from integration to remapping. *Sensory cue integration*, 15:224–250, 2011.

[53] Y. Fan, D. Jiang, M. Wang, and Z. Song. A new markerless patient-to-image registration method using a portable 3d scanner. *Medical physics*, 41(10):101910, 2014.

[54] F. Faure, C. Duriez, H. Delingette, J. Allard, B. Gilles, S. Marchesseau, H. Talbot, H. Courtecuisse, G. Bousquet, I. Peterlik, et al. Sofa: A multi-model framework for interactive physical simulation. In *Soft tissue biomechanical modeling for computer assisted surgery*, pages 283–321. Springer, 2012.

[55] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9): 1323–1341, 2012.

[56] M. J. Fischer and E. B. Strong. Ar surgical navigation with surface tracing: comparing in-situ visualization with tool-tracking guidance for neurosurgical applications. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1170–1179. IEEE, 2025.

[57] R. M. Garcia, R. Dhangana, and W. S. Lindquester. Percutaneous thermal and cryoablation of lung masses compared to sub-lobar lung resection: Medicare trends from 2012 to 2021. *Clinical Imaging*, page 110523, 2025.

[58] V. García-Vázquez, F. Von Haxthausen, S. Jäckle, C. Schumann, I. Kuhlemann, J. Bouchagiar, A.-C. Höfer, F. Matysiak, G. Hüttmann, J. P. Goltz, et al. Navigation and visualisation with hololens in endovascular aortic repair. *Innovative surgical sciences*, 3(3):167–177, 2018.

[59] A. GATTI. Identification of the stress-free configuration of patient-specific and

anatomically complex vascular models: a new strategy to stabilize the sellier's method. 2019.

[60] W. W. Gaver. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology*, 5(1):1–29, 1993.

[61] N. Gerber, K. A. Gavaghan, B. J. Bell, T. M. Williamson, C. Weisstanner, M.-D. Caversaccio, and S. Weber. High-accuracy patient-to-image registration for the facilitation of image-guided robotic microsurgery on the head. *IEEE Transactions on Biomedical Engineering*, 60(4):960–968, 2013.

[62] L. Geronzi, A. Bel-Brunon, A. Martinez, M. Rochette, M. Sensale, O. Bouchot, A. Lalande, S. Lin, P. P. Valentini, and M. E. Biancolini. Calibration of the mechanical boundary conditions for a patient-specific thoracic aorta model including the heart motion effect. *IEEE Transactions on Biomedical Engineering*, 70(11): 3248–3259, 2023.

[63] M. Giménez, B. Gallix, G. Costamagna, J.-N. Vauthey, M. Moche, G. Wakabayashi, R. Bale, L. Swanström, J. Futterer, D. Geller, et al. Definitions of computer-assisted surgery and intervention, image-guided surgery and intervention, hybrid operating room, and guidance systems: Strasbourg international consensus study. *Annals of Surgery Open*, 1(2):e021, 2020.

[64] J. Gindre, A. Bel-Brunon, A. Kaladji, A. Duménil, M. Rochette, A. Lucas, P. Haigron, and A. Combescure. Finite element simulation of the insertion of guidewires during an evar procedure: example of a complex patient case, a first step toward patient-specific parameterized models. *International journal for numerical methods in biomedical engineering*, 31(7):e02716, 2015.

[65] J. Gindre, A. Bel-Brunon, M. Rochette, A. Lucas, A. Kaladji, P. Haigron, and A. Combescure. Patient-specific finite-element simulation of the insertion of guidewire during an evar procedure: guidewire position prediction validation on 28 cases. *IEEE Transactions on Biomedical Engineering*, 64(5):1057–1066, 2016.

[66] L. Gionfrida and A. Roginska. A novel sonification approach to support the diagnosis of alzheimer's dementia. *Frontiers in neurology*, 8:647, 2017.

[67] C. Gsaxner, J. Li, A. Pepe, D. Schmalstieg, and J. Egger. Inside-out instrument tracking for surgical navigation in augmented reality. In *Proceedings of the 27th ACM symposium on virtual reality software and technology*, pages 1–11, 2021.

[68] H. Guo, X. Xu, X. Song, S. Xu, H. Chao, J. Myers, B. Turkbey, P. A. Pinto,

B. J. Wood, and P. Yan. Ultrasound frame-to-volume registration via deep learning for interventional guidance. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 70(9):1016–1025, 2022.

[69] X. T. Ha, M. Ourak, O. Al-Ahmad, D. Wu, G. Borghesan, A. Menciassi, and E. Vander Poorten. Robust catheter tracking by fusing electromagnetic tracking, fiber bragg grating and sparse fluoroscopic images. *IEEE Sensors Journal*, 21(20): 23422–23434, 2021.

[70] J. Hajek, M. Unberath, J. Fotouhi, B. Bier, S. C. Lee, G. Osgood, A. Maier, M. Armand, and N. Navab. Closing the calibration loop: an inside-out-tracking paradigm for augmented reality in orthopedic surgery. In *International conference on medical image computing and computer-assisted intervention*, pages 299–306. Springer, 2018.

[71] J. E. Hammers, S. Pirozzi, D. Lindsay, O. Kaidar-Person, X. Tan, R. C. Chen, S. K. Das, and P. Mavroidis. Evaluation of a commercial dir platform for contour propagation in prostate cancer patients treated with imrt/vmat. *Journal of Applied Clinical Medical Physics*, 21(2):14–25, 2020.

[72] Z. Han and Q. Dou. A review on organ deformation modeling approaches for reliable surgical navigation using augmented reality. *Computer Assisted Surgery*, 29(1): 2357164, 2024.

[73] Z. Han, J. Zhou, J. Pei, J. Qin, Y. Fan, and Q. Dou. Towards reliable ar-guided surgical navigation: interactive deformation modeling with data-driven biomechanics and prompts. *IEEE Transactions on Medical Imaging*, 2025.

[74] C. Hansen, D. Black, C. Lange, F. Rieber, W. Lamadé, M. Donati, K. J. Oldhafer, and H. K. Hahn. Auditory support for resection guidance in navigated liver surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9 (1):36–43, 2013.

[75] M. Harders, R. Hutter, A. Rutz, P. Niederert, and G. Székely. Comparing a simplified fem approach with the mass-spring model for surgery simulation. In *Medicine Meets Virtual Reality 11*, pages 103–109. IOS Press, 2003.

[76] J. Harms, E. Schreibmann, N. S. Mccall, M. S. Lloyd, K. A. Higgins, and R. Castillo. Cardiac motion and its dosimetric impact during radioablation for refractory ventricular tachycardia. *Journal of applied clinical medical physics*, 24(6):e13925, 2023.

[77] B. He, S. Zhao, Y. Dai, J. Wu, H. Luo, J. Guo, Z. Ni, T. Wu, F. Kuang, H. Jiang,

et al. A robust and automatic ct-3d ultrasound registration method based on segmentation, context, and edge hybrid metric. *Medical Physics*, 50(10):6243–6258, 2023.

[78] H. Heibel, B. Glocker, M. Groher, M. Pfister, and N. Navab. Interventional tool tracking using discrete optimization. *IEEE transactions on medical imaging*, 32(3): 544–555, 2012.

[79] T. Hermann and A. Hunt. Parameter mapping sonification. In T. Hermann, A. Hunt, and J. G. Neuhoff, editors, *The Sonification Handbook*, volume 1, chapter 15, pages 273–297. Logos Verlag, Berlin, 2011. URL `https://sonification.de/handbook`.

[80] T. Hermann and M. Weger. Data-driven auditory contrast enhancement for everyday sounds and sonifications. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, 2019.

[81] T. Hermann, T. Nattkemper, W. Schubert, and H. Ritter. Sonification of multichannel image data. *Proc. of the Mathematical and Engineering Techniques in Medical and Biological Sciences (METMBS 2000)*, pages 745–750, 2000.

[82] T. Hermann, A. Hunt, J. G. Neuhoff, et al. *The sonification handbook*, volume 1. Logos Verlag Berlin, 2011.

[83] C. L. Hewitson, M. J. Crossley, J. Cartmill, and D. M. Kaplan. Sensorimotor challenges in minimally invasive surgery: A theoretically-oriented review. *Human Factors*, 67(2):141–165, 2025.

[84] D. Hinck, M. Segeroth, J. Miazza, D. Berdajs, J. Bremerich, J. Wasserthal, and M. Pradella. Automatic segmentation of cardiovascular structures on chest ct data sets: an update of the totalsegmentator. *European journal of radiology*, 185:112006, 2025.

[85] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[86] J. C. Hsu, D. Darden, B. M. Glover, B. J. Colley III, C. Steinberg, B. Thibault, C. Jewell, M. Bernard, P. B. Tabereaux, U. Siddiqui, et al. Performance and acute procedural outcomes of the ensite precision™ cardiac mapping system for electrophysiology mapping and ablation procedures: results from the ensite precision™ observational study. *Journal of Interventional Cardiac Electrophysiology*, 65(1): 141–151, 2022.

[87] Y. Hu, S. Du, T. Xu, and Y. Lei. A novel method for estimating friction boundary conditions and elastic parameters of soft tissue materials. *Journal of Tribology*, pages 1–48, 2026.

[88] C. M. Hults, Y. Ding, G. G. Xie, R. Raja, W. Johnson, A. Lee, and D. J. Simons. Inattentional blindness in medicine. *Cognitive Research: Principles and Implications*, 9(1):18, 2024.

[89] A. Illanes, A. Boese, I. Maldonado, A. Pashazadeh, A. Schaufler, N. Navab, and M. Friebe. Novel clinical device tracking and tissue event characterization using proximally placed audio signal acquisition and processing. *Scientific reports*, 8(1): 12070, 2018.

[90] H. Iqbal and F. R. y Baena. Semi-automatic infrared calibration for augmented reality systems in surgery. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4957–4964. IEEE, 2022.

[91] Itseez. Open source computer vision library. `https://github.com/itseez/opencv`, 2015.

[92] J. Jiang and T. J. Hall. A parallelizable real-time motion tracking algorithm with applications to ultrasonic strain imaging. *Physics in Medicine & Biology*, 52(13): 3773, 2007.

[93] D. W. Jones, L. Stangenberg, N. J. Swerdlow, M. Alef, R. Lo, F. Shuja, and M. L. Schermerhorn. Image fusion and 3-dimensional roadmapping in endovascular surgery. *Annals of Vascular Surgery*, 52:302–311, 2018.

[94] C. Jung, G. Wolff, B. Wernly, R. R. Bruno, M. Franz, P. C. Schulze, J. N. A. Silva, J. R. Silva, D. L. Bhatt, and M. Kelm. Virtual and augmented reality in cardiovascular care: state-of-the-art and future perspectives. *Cardiovascular Imaging*, 15(3): 519–532, 2022.

[95] N. Kadoya, Y. Fujita, Y. Katsuta, S. Dobashi, K. Takeda, K. Kishi, M. Kubozono, R. Umezawa, T. Sugawara, H. Matsushita, et al. Evaluation of various deformable image registration algorithms for thoracic images. *Journal of radiation research*, 55 (1):175–182, 2014.

[96] J. Kalkhof, C. González, and A. Mukhopadhyay. Med-NCA: Robust and lightweight segmentation with neural cellular automata. In *Information Processing in Medical Imaging (IPMI)*, pages 705–716. Springer, 2023.

[97] J. Kalkhof et al. Temporal neural cellular automata: Application to modeling of

contrast enhancement in breast MRI. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2025.

[98] P. R. Kantan, S. Dahl, and E. G. Spaich. Sound-guided 2-d navigation: Effects of information concurrency and coordinate system. In *Nordic Human-Computer Interaction Conference*, pages 1–11, 2022.

[99] C. A. Kemper, I.-F. Talos, A. Golby, P. M. Black, R. Kikinis, W. E. L. Grimson, and S. K. Warfield. An anisotropic material model for image guided neurosurgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 267–275. Springer, 2004.

[100] J. Kim, B. Peruski, C. Hunley, S. Kwon, and S. Baek. Influence of surrounding tissues on biomechanics of aortic wall. *International journal of experimental and computational biomechanics*, 2(2):105–117, 2013.

[101] S. Kim, H. Kim, and Y. Jung. An automated marker-less registration approach using neural radiance fields for potential use in mixed reality-based computer-aided surgical navigation of paranasal sinus. *Computers*, 14(1):5, 2024.

[102] D. Kirubakaran. Review on emerging frontiers in medical sensing: Innovations shaping the future of healthcare. *Biomedical Materials & Devices*, pages 1–21, 2025.

[103] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009.

[104] J. S. Koruth, A. Aryana, S. R. Dukkipati, H.-N. Pak, Y.-H. Kim, E. A. Sosa, M. Scanavacca, S. Mahapatra, G. Ailawadi, V. Y. Reddy, et al. Unusual complications of percutaneous epicardial access and epicardial mapping and ablation of cardiac arrhythmias. *Circulation: Arrhythmia and Electrophysiology*, 4(6):882–888, 2011.

[105] F. Kral, E. J. Puschban, H. Riechelmann, and W. Freysinger. Comparison of optical and electromagnetic tracking for navigated lateral skull base surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 9(2):247–252, 2013.

[106] J. Krücker, S. Xu, N. Glossop, A. Viswanathan, J. Borgert, H. Schulz, and B. J. Wood. Electromagnetic tracking for thermal ablation and biopsy guidance: clinical

evaluation of spatial accuracy. *Journal of Vascular and Interventional Radiology*, 18(9):1141–1150, 2007.

[107] H. J. Krumb, J. Mehtali, J. Verde, A. Mukhopadhyay, and C. Essert. Cryotrack: Planning and navigation for computer assisted cryoablation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 99–108. Springer, 2024.

[108] U. Kühnapfel, H. K. Cakmak, and H. Maaß. Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers & graphics*, 24(5):671–682, 2000.

[109] N. B. Kumsa, T.-L. Kelly, E. E. Roughead, R. Tavella, and M. H. Gillam. Temporal trends in percutaneous coronary intervention in australia: A retrospective analysis from 2000–2021. *Hellenic Journal of Cardiology*, 80:21–30, 2024.

[110] Y. Kwong, A. O. Mel, G. Wheeler, and J. M. Troupis. F our-dimensional computed tomography (4dct): a review of the current status and applications. *Journal of medical imaging and radiation oncology*, 59(5):545–554, 2015.

[111] F. Lalys and P. Jannin. Surgical process modelling: a review. *International journal of computer assisted radiology and surgery*, 9(3):495–511, 2014.

[112] F. Lareyre, C. Mialhe, B. Nasr, E. Poggi, G. Di Lorenzo, K. Rajhi, A. Chaudhuri, and J. Raffort. Extended and augmented reality in vascular surgery: Opportunities and challenges. In *Seminars in Vascular Surgery*, volume 37, pages 321–325. Elsevier, 2024.

[113] K. Latifi, J. Caudell, G. Zhang, D. Hunt, E. G. Moros, and V. Feygelman. Practical quantification of image registration accuracy following the aapm tg-132 report framework. *Journal of applied clinical medical physics*, 19(4):125–133, 2018.

[114] R. Lee, B. Kang, D. Kim, and D. Kim. Correlation based dynamic time warping for ecg waveform. *Applied Sciences*, 16(5):2369, 2026.

[115] L. Lei, J. Zhou, J. Pei, B. Zhao, Y. Jin, Y.-C. J. Teoh, J. Qin, and P.-A. Heng. Epicardium prompt-guided real-time cardiac ultrasound frame-to-volume registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 618–628. Springer, 2024.

[116] G. Lemaitre, O. Houix, Y. Visell, K. Franinović, N. Misdariis, and P. Susini. Toward the design and evaluation of continuous sound in tangible interfaces: The spinotron. *International Journal of Human-Computer Studies*, 67(11):976–993, 2009.

[117] J. Li, Z. Deng, N. Shen, Z. He, L. Feng, Y. Li, and J. Yao. A fully automatic surgical registration method for percutaneous abdominal puncture surgical navigation. *Computers in Biology and Medicine*, 136:104663, 2021.

[118] R. Liao, L. Zhang, Y. Sun, S. Miao, and C. Chefd'Hotel. A review of recent advances in registration techniques applied to minimally invasive therapy. *IEEE transactions on multimedia*, 15(5):983–1000, 2013.

[119] F. Liebmann, S. Roner, M. von Atzigen, D. Scaramuzza, R. Sutter, J. Snedeker, M. Farshad, and P. Fürnstahl. Pedicle screw navigation using surface digitization on the microsoft hololens. *International journal of computer assisted radiology and surgery*, 14(7):1157–1165, 2019.

[120] Q. Lin, X. Guo, Y. Xie, K. Peng, R. Yang, and K. Cai. Surface matching-based markerless global optimization registration for improved optical surgical systems in internet of things-enabled operating rooms. *IEEE Transactions on Consumer Electronics*, 70(1):939–946, 2023.

[121] J. Liu, S. J. Al'Aref, G. Singh, A. Caprio, A. A. A. Moghadam, S.-J. Jang, S. C. Wong, J. K. Min, S. Dunham, and B. Mosadegh. An augmented reality system for image guidance of transcatheter procedures for structural heart disease. *PloS one*, 14(7):e0219174, 2019.

[122] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998.

[123] G. Luraghi, F. Migliavacca, A. García-González, C. Chiastra, A. Rossi, D. Cao, G. Stefanini, and J. F. Rodriguez Matas. On the modeling of patient-specific transcatheter aortic valve replacement: A fluid–structure interaction approach: Luraghi et al. *Cardiovascular engineering and technology*, 10(3):437–455, 2019.

[124] H. Ma, I. Smal, J. Daemen, and T. van Walsum. Dynamic coronary roadmapping via catheter tip tracking in x-ray fluoroscopy with deep learning based bayesian filtering. *Medical image analysis*, 61:101634, 2020.

[125] L. Ma, T. Huang, J. Wang, and H. Liao. Visualization, registration and tracking techniques for augmented reality guided surgery: a review. *Physics in Medicine & Biology*, 68(4):04TR02, 2023.

[126] M. Mach, S. Okutucu, T. Kerbel, A. Arjomand, S. G. Fatihoglu, P. Werner, P. Si-

mon, and M. Andreas. Vascular complications in tavr: incidence, clinical impact, and management. *Journal of clinical medicine*, 10(21):5046, 2021.

[127] A. Mack. Inattentional blindness: Looking without seeing. *Current directions in psychological science*, 12(5):180–184, 2003.

[128] M. J. Mack, M. B. Leon, V. H. Thourani, R. Makkar, S. K. Kodali, M. Russo, S. R. Kapadia, S. C. Malaisrie, D. J. Cohen, P. Pibarot, et al. Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. *New England Journal of Medicine*, 380(18):1695–1705, 2019.

[129] U. Magana-Salgado, P. Namburi, M. Feigin-Almon, R. Pallares-Lopez, and B. Anthony. A comparison of point-tracking algorithms in ultrasound videos from the upper limb. *Biomedical Engineering Online*, 22(1):52, 2023.

[130] S. Malhotra, O. Halabi, S. P. Dakua, J. Padhan, S. Paul, and W. Palliyali. Augmented reality in surgical navigation: a review of evaluation and validation metrics. *Applied Sciences*, 13(3):1629, 2023.

[131] A. Mangla and S. Gupta. Vascular complications post-transcatheter aortic valve procedures. *Indian Heart Journal*, 68(5):724–731, 2016.

[132] P. Mascagni, D. Alapatt, L. Sestini, M. S. Altieri, A. Madani, Y. Watanabe, A. Alseidi, J. A. Redan, S. Alfieri, G. Costamagna, et al. Computer vision in surgery: from potential to clinical value. *npj Digital Medicine*, 5(1):163, 2022.

[133] A. Mason, R. Paulsen, J. M. Babuska, S. Rajpal, S. Burneikiene, E. L. Nelson, and A. T. Villavicencio. The accuracy of pedicle screw placement using intraoperative image guidance systems: a systematic review. *Journal of Neurosurgery: Spine*, 20 (2):196–203, 2014.

[134] S. Mathew and et Al. Epicardial access for vt ablation: analysis of two different puncture techniques, incidence of adhesions and complication management. *Clinical Research in Cardiology*, 110:810–821, 2021.

[135] S. Matinfar. *From Tissue to Sound: Medical Extended Reality (XR) Paradigms for Enhanced Surgical Decision-Making and Action*. PhD thesis, Technische Universität München, 2025.

[136] S. Matinfar, M. A. Nasseri, U. Eck, H. Roodaki, N. Navab, C. P. Lohmann, M. Maier, and N. Navab. Surgical soundtracks: Towards automatic musical augmentation of surgical procedures. In *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC,*

Canada, September 11-13, 2017, Proceedings, Part II 20, pages 673–681. Springer, 2017.

[137] S. Matinfar, M. A. Nasseri, U. Eck, M. Kowalsky, H. Roodaki, N. Navab, C. P. Lohmann, M. Maier, and N. Navab. Surgical soundtracks: Automatic acoustic augmentation of surgical procedures. *International journal of computer assisted radiology and surgery*, 13:1345–1355, 2018.

[138] S. Matinfar, M. Salehi, D. Suter, M. Seibold, S. Dehghani, N. Navab, F. Wanivenhaus, P. Fürnstahl, M. Farshad, and N. Navab. Sonification as a reliable alternative to conventional visual surgical navigation. *Scientific Reports*, 13(1):5930, 2023.

[139] S. Matinfar, S. Dehghani, M. Sommersperger, and N. Navab. Ocular stethoscope: Auditory support for retinal membrane peeling. In *accepted in International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024.

[140] S. Matinfar, S. Dehghani, M. Salehi, M. Sommersperger, N. Navab, K. Faridpooya, M. Fairhurst, and N. Navab. From tissue to sound: A new paradigm for medical sonic interaction design. *Medical Image Analysis*, 103:103571, 2025.

[141] J. Mehtali, J. M. Verde, and C. Essert. C-NCA: Chained neural cellular automata for fast and accurate thermal ablation estimation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 67–77. Springer, 2025.

[142] A. Mendizabal, A. Odot, and S. Cotin. Deep learning for real-time computational biomechanics. In *Reduced order models for the biomechanics of living organs*, pages 95–126. Elsevier, 2023.

[143] A. Mendizabal, E. Tagliabue, and D. Dall'Alba. Intraoperative estimation of liver boundary conditions from multiple partial surfaces. *International Journal of Computer Assisted Radiology and Surgery*, 18(7):1295–1302, 2023.

[144] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[145] F. Milletari, V. Belagiannis, N. Navab, and P. Fallavollita. Fully automatic catheter localization in c-arm images using l1-sparse coding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 570–577. Springer, 2014.

[146] H. N. Modi, S. W. Suh, H. Fernandez, J. H. Yang, and H.-R. Song. Accuracy

and safety of pedicle screw placement in neuromuscular scoliosis with free-hand technique. *European Spine Journal*, 17(12):1686–1696, 2008.

[147] M. Moerel, F. De Martino, and E. Formisano. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, 32(41):14205–14216, 2012.

[148] B. C. Moore. *An introduction to the psychology of hearing*. Brill, 2012.

[149] A. Mordvintsev, E. Randazzo, E. Niklasson, and M. Levin. Growing neural cellular automata. *Distill*, 5(2):e23, 2020.

[150] R. Moreta-Martínez, I. Rubio-Pérez, M. García-Sevilla, L. García-Elcano, and J. Pascau. Evaluation of optical tracking and augmented reality for needle navigation in sacral nerve stimulation. *Computer methods and programs in biomedicine*, 224:106991, 2022.

[151] C. Moro, J. Birt, Z. Stromberga, C. Phelps, J. Clark, P. Glasziou, and A. M. Scott. Virtual and augmented reality enhancements to medical and science student physiology and anatomy test performance: A systematic review and meta-analysis. *Anatomical sciences education*, 14(3):368–376, 2021.

[152] T. A. Mulder, T. van de Velde, E. Dokter, B. Boekestijn, T. J. Olgers, M. P. Bauer, and B. P. Hierck. Unravelling the skillset of point-of-care ultrasound: a systematic review. *The Ultrasound Journal*, 15(1):19, 2023.

[153] F. Müller, S. Roner, F. Liebmann, J. M. Spirig, P. Fürnstahl, and M. Farshad. Augmented reality navigation for spinal pedicle screw instrumentation using intraoperative 3d imaging. *The Spine Journal*, 20(4):621–628, 2020.

[154] R. Munafò, S. Saitta, D. Tondi, G. Ingallina, P. Denti, F. Maisano, E. Agricola, and E. Votta. Automatic 4d mitral valve segmentation from transesophageal echocardiography: a semi-supervised learning approach. *Medical & Biological Engineering & Computing*, pages 1–16, 2025.

[155] R. Munafò, S. Saitta, L. Vicentini, D. Tondi, V. Ruozzi, F. Sturla, G. Ingallina, A. Guidotti, E. Agricola, and E. Votta. Mitraclip device automated localization in 3d transesophageal echocardiography via deep learning. *Computer Methods and Programs in Biomedicine*, page 109083, 2025.

[156] A. Murphy and P. O'Shea. Cardiac gating (CT). `https://radiopaedia.org/articles/40208`. Radiopaedia.

[157] N. Navab, A. Martin-Gomez, M. Seibold, M. Sommersperger, T. Song, A. Winkler, K. Yu, and U. Eck. Medical augmented reality: definition, principle components, domain modeling, and design-development-validation process. *Journal of Imaging*, 9(1):4, 2022.

[158] E. Niklasson, A. Mordvintsev, E. Randazzo, and M. Levin. Self-organising textures. *Distill*, 6(2):e00027.003, 2021.

[159] A. I. Omara, M. Wang, Y. Fan, and Z. Song. Anatomical landmarks for point-matching registration in image-guided neurosurgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 10(1):55–64, 2014.

[160] D. Ostler-Mildner, L. Wegener, J. Fuchtmann, H. Feussner, D. Wilhelm, and N. Navab. The sound of surgery-development of an acoustic trocar system enabling laparoscopic sound analysis. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–9, 2024.

[161] Z. Ouyang, P. Schoenhagen, O. Wazni, P. Tchou, W. I. Saliba, J. H. Suh, and P. Xia. Analysis of cardiac motion without respiratory motion for cardiac stereotactic body radiation therapy. *Journal of Applied Clinical Medical Physics*, 21(10):48–55, 2020.

[162] K. O'donoghue, H. A. Jaeger, and P. Cantillon-Murphy. A radiolucent electromagnetic tracking system for use with intraoperative x-ray imaging. *Sensors*, 21(10): 3357, 2021.

[163] M. C. Palumbo. Unveiling the invisible: mixed reality based approaches for surgical empowerment. 2022.

[164] M. C. Palumbo, S. Saitta, M. Schiariti, M. C. Sbarra, E. Turconi, G. Raccuia, J. Fu, V. Dallolio, P. Ferroli, E. Votta, et al. Mixed reality and deep learning for external ventricular drainage placement: A fast and automatic workflow for emergency treatments. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 147–156. Springer, 2022.

[165] S. Pasta, V. Agnese, A. Gallo, F. Cosentino, M. Di Giuseppe, G. Gentile, G. M. Raffa, J. F. Maalouf, H. I. Michelena, D. Bellavia, et al. Shear stress and aortic strain associations with biomarkers of ascending thoracic aortic aneurysm. *The Annals of thoracic surgery*, 110(5):1595–1604, 2020.

[166] M. Pescio, C. Li, D. Kundrat, M. Casadio, and G. Dagnino. A simulation environment for robot-assisted endovascular interventions. *International Journal of Computer Assisted Radiology and Surgery*, 20(11):2259–2267, 2025.

[167] I. Peterlík, C. Duriez, and S. Cotin. Modeling and real-time simulation of a vascularized liver tissue. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 50–57. Springer, 2012.

[168] C. Pohl, N. N. N. Nazirun, N. Hamzah, and S. S. Tamin. Multimodal medical image fusion in cardiovascular applications. In *Medical Imaging Technology: Reviews and Computational Applications*, pages 91–109. Springer, 2015.

[169] D.-L. Popa, M.-L. Mocanu, R.-T. Popa, L.-F. Bărbulescu, L. N. Bărbulescu, and C. Ciobîrcă. Augmented and mixed reality for collaborative cardiac planning using 3d and 4d medical imaging data. In *2020 24th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 384–389. IEEE, 2020.

[170] Z. Qi, M. H. Bopp, C. Nimsky, X. Chen, X. Xu, Q. Wang, Z. Gan, S. Zhang, J. Wang, H. Jin, et al. A novel registration method for a mixed reality navigation system based on a laser crosshair simulator: A technical note. *Bioengineering*, 10 (11):1290, 2023.

[171] M. Raad and G. E. Supple. Epicardial ventricular tachycardia ablation: patient selection, access, and ablation techniques and strategies to manage complications. *Clinical Electrophysiology*, 10(1):142–164, 2024.

[172] A. Ramadani, M. Bui, T. Wendler, H. Schunkert, P. Ewert, and N. Navab. A survey of catheter tracking concepts and methodologies. *Medical image analysis*, 82:102584, 2022.

[173] A. R. Rao. A spatio-temporal model of multi-sensory learning that demonstrates improved object recall. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 5043–5050. IEEE, 2016.

[174] H. Rasheed, R. Dorent, M. Fehrentz, T. Kapur, W. M. Wells III, A. Golby, S. Frisken, J. A. Schnabel, and N. Haouchine. Learning to match 2d keypoints across preoperative mr and intraoperative ultrasound. In *International Workshop on Advances in Simplifying Medical Ultrasound*, pages 78–87. Springer, 2024.

[175] I. E. Rassi and J.-M. E. Rassi. A review of haptic feedback in tele-operated robotic surgery. *Journal of Medical Engineering & Technology*, 44:247–254, 7 2020. ISSN 0309-1902. doi: 10.1080/03091902.2020.1772391. haptic review.

[176] D. Reichard, D. Häntsch, S. Bodenstedt, S. Suwelack, M. Wagner, H. Kenngott, B. Müller-Stich, L. Maier-Hein, R. Dillmann, and S. Speidel. Projective biomechan-

ical depth matching for soft tissue registration in laparoscopic surgery. *International journal of computer assisted radiology and surgery*, 12(7):1101–1110, 2017.

[177] S. J. Rhee, S. H. Park, H. M. Cho, and J. T. Suh. Comparison of precision between optical and electromagnetic navigation systems in total knee arthroplasty. *Knee surgery & related research*, 26(4):214, 2014.

[178] H. Rivaz, E. M. Boctor, M. A. Choti, and G. D. Hager. Real-time regularized ultrasound elastography. *IEEE Transactions on Medical Imaging*, 30(4):928–945, 2011.

[179] J. Romero and et Al. Epicardial access complications during electrophysiology procedures. *Journal of cardiovascular electrophysiology*, 32(7):1985–1994, 2021.

[180] J. Romero, K. Shivkumar, L. Di Biase, R. Avendano, R. D. Anderson, A. Natale, and S. Kumar. Mastering the art of epicardial access in cardiac electrophysiology. *Heart Rhythm*, 16(11):1738–1749, 2019.

[181] W. J. Rucklidge. Efficiently locating objects using the hausdorff distance. *International Journal of computer vision*, 24(3):251–270, 1997.

[182] L. Schütz, S. Matinfar, N. Navab, M. Fairhurst, U. Eck, and N. Navab. A framework for multimodal medical image interaction. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[183] D. Schwarz. Corpus-based concatenative synthesis. *IEEE signal processing magazine*, 24(2):92–104, 2007.

[184] D. Schwarz and B. Hackbarth. Navigating variation: composing for audio mosaicing. In *International Computer Music Conference (ICMC)*, pages 1–1, 2012.

[185] D. Scorza, S. El Hadji, C. Cortes, A. Bertelsen, F. Cardinale, G. Baselli, C. Essert, and E. De Momi. Surgical planning assistance in keyhole and percutaneous surgery: a systematic review. *Medical image analysis*, 67:101820, 2021.

[186] J. Seetohul, M. Shafiee, and K. Sirlantzis. Augmented reality (ar) for surgical robotic and autonomous systems: state of the art, challenges, and solutions. *Sensors*, 23 (13):6202, 2023.

[187] M. Selim, D. Dresscher, and M. Abayazid. Virtual needle insertion with enhanced haptic feedback for guidance and needle–tissue interaction forces. *Sensors*, 24(17): 5560, 2024.

[188] L. Shi, Q. Chen, S. Barley, Y. Cui, L. Shang, J. Qiu, and Y. Rong. Benchmarking

of deformable image registration for multiple anatomic sites using digital data sets with ground-truth deformation vector fields. *Practical Radiation Oncology*, 11(5): 404–414, 2021.

[189] T. E. Shkurti and M. C. Çavuşoğlu. A systematic review of task automation in surgical robotics. *IEEE Transactions on Medical Robotics and Bionics*, 2025.

[190] G. Singh and A. Chanda. Mechanical properties of whole-body soft human tissues: a review. *Biomedical Materials*, 16(6):062004, 2021.

[191] J. O. Smith. Physical modeling using digital waveguides. *Computer music journal*, 16(4):74–91, 1992.

[192] A. Sorriento, M. B. Porfido, S. Mazzoleni, G. Calvosa, M. Tenucci, G. Ciuti, and P. Dario. Optical and electromagnetic tracking systems for biomedical applications: A critical review on potentialities and limitations. *IEEE reviews in biomedical engineering*, 13:212–232, 2019.

[193] E. Sosa, M. Scanavacca, A. D'avila, and E. PILLEGGI. A new technique to perform epicardial mapping in the electrophysiology laboratory. *Journal of cardiovascular electrophysiology*, 7(6):531–536, 1996.

[194] E. Sosa, M. Scanavacca, A. D'AVILA, J. Piccioni, O. Sanchez, J. L. Velarde, M. Silva, and B. Reolao. Endocardial and epicardial ablation guided by nonsurgical transthoracic epicardial mapping to treat recurrent ventricular tachycardia. *Journal of cardiovascular electrophysiology*, 9(3):229–239, 1998.

[195] J. Sra. Cardiac image registration. *Journal of Atrial Fibrillation*, 1(3):25, 2008.

[196] C. Stapleton and J. Davies. Imagination: The third reality to the virtuality continuum. In *2011 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities*, pages 53–60. IEEE, 2011.

[197] C. Stapleton, C. Hughes, M. Moshell, P. Micikevicius, and M. Altman. Applying mixed reality to entertainment. *Computer*, 35(12):122–124, 2002.

[198] K. Steeg, G. A. Krombach, and M. H. Friebe. A review of needle navigation technologies in minimally invasive cardiovascular surgeries—toward a more effective and easy-to-apply process. *Diagnostics*, 15(2):197, 2025.

[199] B. E. Stein. *The new handbook of multisensory processing*. Mit Press, 2012.

[200] S. Sundaram, P. Tung, D. Yoo, D. Lakkireddy, U. Siddiqui, J. Osorio, K. Nilsson, D. Shah, P. Sommer, and M. Lo. Safety and effectiveness of zero fluoroscopy proce-

dures using a novel dual-energy ablation catheter. *Journal of Interventional Cardiac Electrophysiology*, 68(9):1927–1929, 2025.

[201] P. Susini, N. Misdariis, G. Lemaitre, and O. Houix. Naturalness influences the perceived usability and pleasantness of an interface's sonic feedback. *Journal on Multimodal User Interfaces*, 5:175–186, 2012.

[202] M. Swale, S. Mikell, J. Gard, T. M. Munger, S. J. Asirvatham, and P. A. Friedman. Epicardial access: patient selection, anatomy, and a stepwise approach. *J Innov Card Rhythm Manag*, 2:239–249, 2011.

[203] V. Tacher, A. Petit, H. Derbel, L. Novelli, M. Vitellius, F. Ridouani, A. Luciani, A. Rahmouni, C. Duvoux, C. Salloum, et al. Three-dimensional image fusion guidance for transjugular intrahepatic portosystemic shunt placement. *Cardiovascular and Interventional Radiology*, 40(11):1732–1739, 2017.

[204] E. Tagliabue, D. Dall'Alba, E. Magnabosco, I. Peterlik, and P. Fiorini. Biomechanical modelling of probe to tissue interaction during ultrasound scanning. *International Journal of Computer Assisted Radiology and Surgery*, 15(8):1379–1387, 2020.

[205] E. Tagliabue, M. Piccinelli, D. Dall'Alba, J. Verde, M. Pfeiffer, R. Marin, S. Speidel, P. Fiorini, and S. Cotin. Intra-operative update of boundary conditions for patient-specific surgical simulation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 373–382. Springer, 2021.

[206] Y. Takahashi and C. Tanaka-Esposito. Tactiflex™ ablation catheter, sensor enabled™: Initial experience from the tactiflex af ide trial and in japan after early market release. *J. Atr. Fibrillation*, 16:22–25, 2023.

[207] A. K. Z. Tehrani and H. Rivaz. Infusing physically inspired known operators in deep models of ultrasound elastography. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Springer, 2023.

[208] F. E. Theunissen and J. E. Elie. Neural processing of natural sounds. *Nature Reviews Neuroscience*, 15(6):355–366, 2014.

[209] Y. Tillier, A. Paccini, M. Durand-Reville, and J.-L. Chenot. Finite element modeling for soft tissue surgery based on linear and nonlinear elasticity behavior. *Computer Aided Surgery*, 11(2):63–68, 2006.

[210] Y. Tong, Y. Yin, J. Lu, T. Liu, J. Chen, P. Cheng, and G. Gong. Quantification of heart, pericardium, and left ventricular myocardium movements during the cardiac

cycle for thoracic tumor radiotherapy. *OncoTargets and therapy*, pages 547–554, 2018.

[211] P. Tu, P. Hu, J. Wang, and X. Chen. From coarse to fine: Non-rigid sparse-dense registration for deformation-aware liver surgical navigation. *IEEE Transactions on Biomedical Engineering*, 71(9):2663–2677, 2024.

[212] R. Urrutia, F. A. Abdelfattah, A. Boese, C. Hansen, and A. Illanes. Needle puncture detection using vibroacoustic sensing in layered phantoms. In *Current Directions in Biomedical Engineering*, volume 11, pages 112–115. De Gruyter, 2025.

[213] U.S. Food and Drug Administration. 510(k) summary for CorPath GRX System (k202275), Dec. 2020. URL `http://510k.innolitics.com/device/K202275`. FDA Premarket Notification Database.

[214] C. Varghese, E. M. Harrison, G. O'Grady, and E. J. Topol. Artificial intelligence in surgery. *Nature medicine*, 30(5):1257–1268, 2024.

[215] J. Villeneuve and J. Leonard. Mass-interaction physical models for sound and multi-sensory creation: Starting anew. In *Proceedings of the 16th Sound & Music Computing Conference*, pages 187–194, 2019.

[216] F. Von Haxthausen, R. Moreta-Martinez, A. Pose Díez de la Lastra, J. Pascau, and F. Ernst. Ultrarsound: in situ visualization of live ultrasound images using hololens 2. *International journal of computer assisted radiology and surgery*, 17(11): 2081–2091, 2022.

[217] E. Wang, Y. Liu, P. Tu, Z. A. Taylor, and X. Chen. Video-based soft tissue deformation tracking for laparoscopic augmented reality-based navigation in kidney surgery. *IEEE Transactions on Medical Imaging*, 43(12):4161–4173, 2024.

[218] H. Wang and Y. Wang. Eureg: End-to-end framework for efficient 2d-3d ultrasound registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 175–185. Springer, 2025.

[219] M. Wang and Z. Song. Automatic localization of the center of fiducial markers in 3d ct/mri images for image-guided neurosurgery. *Pattern Recognition Letters*, 30 (4):414–420, 2009.

[220] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5 (5):e230024, 2023.

[221] M. Weger, T. Hermann, and R. Höldrich. Real-time auditory contrast enhancement. In *Proceedings of the 25th International Conference on Auditory Display (ICAD 2019)*, 2019.

[222] Y. Wei, S. Cotin, J. Dequidt, C. Duriez, J. Allard, E. Kerrien, et al. A (near) real-time simulation method of aneurysm coil embolization. *Aneurysm*, 8(29):223–248, 2012.

[223] E. P. Westebring-van der Putten, R. H. Goossens, J. J. Jakimowicz, and J. Dankelman. Haptics in minimally invasive surgery–a review. *Minimally Invasive Therapy & Allied Technologies*, 17(1):3–16, 2008.

[224] M. Wright, A. Freed, and A. Momeni. 2003: Opensound control: State of the art 2003. *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*, pages 125–145, 2017.

[225] D. Wu, Z. Li, M. H. D. Ansari, X. T. Ha, M. Ourak, J. Dankelman, A. Menciassi, E. De Momi, and E. Vander Poorten. Comparative analysis of interactive modalities for intuitive endovascular interventions. *IEEE Transactions on Visualization and Computer Graphics*, 31(2):1371–1388, 2024.

[226] Y. Ying, X. Fang, Y. Zhao, X. Zhao, Y. Zhou, G. Du, Y. Zhan, T. Gao, A. Li, D. Sun, et al. Sam-myonet: a fine-grained perception myocardial ultrasound segmentation network based on segment anything model with prior knowledge driven. *Biomedical Signal Processing and Control*, 110:108117, 2025.

[227] M. N. Young, S. Kearing, M. A. Albaghdadi, A. Latib, and A. Iribarne. Trends in transcatheter vs surgical mitral valve repair among medicare beneficiaries, 2012 to 2019. *JAMA cardiology*, 7(7):770–772, 2022.

[228] D. Zerbato, S. Galvan, and P. Fiorini. Calibration of mass spring models for organ simulations. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 370–375. IEEE, 2007.

[229] J. Zhang, V. Lu, and V. Khanduja. The impact of extended reality on surgery: a scoping review. *International orthopaedics*, 47(3):611–621, 2023.

[230] Y. Zhang, Z. Yuan, Y. Ding, J. Zhao, Z. Duan, and M. Sun. Real time simulation of tissue cutting based on gpu and cuda for surgical training. In *2010 International Conference on Biomedical Engineering and Computer Science*, pages 1–4. IEEE, 2010.

[231] Z. Zheng, J. Fan, L. Shao, X. Zhang, H. Song, D. Ai, T. Fu, D. Xiao, T. Zhang,

Y. Wang, et al. Simultaneous reconstruction and tracking for motion correction in image guided surgery. *IEEE Transactions on Biomedical Engineering*, 2025.

[232] Y. Zhou, Q. Zuo, N. Chen, L. Zhou, B. Yang, Z. Liu, Y. Liu, L. Tang, S. Dong, and Z. Jiang. Transformer based deep learning for digital image correlation. *Optics and Lasers in Engineering*, 184:108568, 2025.

[233] Y.-J. Zhou, X.-L. Xie, X.-H. Zhou, S.-Q. Liu, G.-B. Bian, and Z.-G. Hou. A real-time multifunctional framework for guidewire morphological and positional analysis in interventional x-ray fluoroscopy. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3):657–667, 2020.

[234] T. Ziemer. Three-dimensional sonification as a surgical guidance tool. *Journal on Multimodal User Interfaces*, 17(4):253–262, 2023.

[235] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media, 2013.

A | Appendix A

Table A.1: Mechanical parameters sensitivity analysis: variation of Young's modulus (E) and Poisson's ratio (ν) from baseline configuration ($E = 1.0$ MPa, $\nu = 0.36$). Metrics quantify local displacement deviation ($\Delta\mathbf{U}$) and global shape change (Chamfer Distance, *ChDist*).

E **(MPa)**	ν	$\Delta\mathbf{U}$ **(mm)**	***ChDist*** **(mm)**
Variation of E at fixed $\nu = 0.36$			
0.5	0.36	0.56	0.14
0.7	0.36	0.22	0.06
0.9	0.36	0.07	0.02
1.0	0.36	**baseline**	**baseline**
1.1	0.36	0.06	0.02
1.2	0.36	0.11	0.03
1.5	0.36	0.24	0.07
2.0	0.36	0.43	0.12
Variation of ν at fixed $E = 1.0$ MPa			
1.0	0.30	0.03	0.01
1.0	0.33	0.01	0.00
1.0	0.36	**baseline**	**baseline**
1.0	0.39	0.02	0.01
1.0	0.42	0.03	0.01
1.0	0.45	0.05	0.01

B | Cardiac Dynamic Reconstruction: Technical Background

This appendix provides the theoretical and technical background for the cardiac dynamic reconstruction methods employed in Chapter 5, Section 5.2.1. Specifically, it describes the deformable image registration technique used to propagate pericardial segmentations across all frames of the 4DCT dataset, enabling accurate 3D reconstruction of the pericardium throughout the cardiac cycle.

B.1. 4-Dimensional Computed Tomography Imaging

CT is a diagnostic imaging technique that utilizes ionizing radiation, specifically x-rays, to produce cross-sectional images, known as slices, of the patient's inner anatomy. These images contain data regarding the distribution of x-ray beam attenuation coefficients $\mu(x, y)$ within the object's volume, enabling the calculation of tissue density along the path of the x-ray beam.

4DCT is an advanced type of CT where a series of CT images are acquired over time with high spatial and temporal resolution. Differently from conventional 3DCT scans, where x-rays are employed to capture the static representation of the patient's anatomy at specific time points, 4DCT generates sequential images over intervals of time to reconstruct gray-scale images at multiple time frames. In other words, time is integrated as a fourth dimension to a conventional 3D computed tomography. The output obtained is a cine-image sequence, a dynamic volume dataset that facilitates the observation of physiological motion and the tracking of anatomical deformations, particularly for medical applications such as radiation therapy and respiratory motion management [110].

B.1.1. ECG Gating

Especially when analyzing heart chambers and related structures, or conversely, when minimizing imaging artifacts resulting from cardiac motion, ECG gating emerges as a common technique for obtaining CT scans at precise time points within the cardiac cycle. The ECG R-R interval is treated as a window covering the entire cardiac cycle and is typically divided into discrete phases ranging from 0% to 100%. In this scheme, the first R wave marks the onset of the cardiac cycle at end diastole, representing the 0% phase. The subsequent R wave corresponds to the 100% of the current cycle, simultaneously aligning with the 0% point of the following cycle [44].

Using the ECG signal to control scanning, two different types of ECG gating are available: prospective or retrospective gating. In prospective gating, the images are acquired at specific time frames by the activation of the x-ray source only for the predetermined portion of the R-R interval that has been selected for the scanning. This technique is commonly employed to reduce motion artifacts considering the triggering during diastolic phase (around 70–85% of cardiac cycle), when the least cardiac motion occurs due to ventricle passive filling [156].

On the contrary, retrospective gating involves scanning throughout the entire cardiac cycle or eventually for multiple cardiac cycles. This approach generates cine-sequences that capture heart motion throughout the entire R-R interval, making it a well-established technique for acquiring information about cardiac dynamics and analyzing cardiac motion for various diagnostic purposes, particularly in cases of arrhythmias and valvular dysfunctions [156].

Following acquisition, images are reconstructed to yield 10 to 20 equally spaced frames spanning the complete R-R interval, facilitating the alignment of CT data with the ECG signal [44]. Once a 4DCT dataset is acquired, various imaging techniques and tracking algorithms are available to assess and quantify anatomical changes over time. Image registration, among others, plays a pivotal role in enabling automatic segmentation propagation across multiple frames.

B.2. Deformable Image Registration

Deformable image registration using 4DCT datasets is a valuable technique for propagating anatomical segmentations across multiple time frames. While 4DCT offers a comprehensive perspective on cardiac anatomy by reconstructing volumetric data over time, image registration serves as a method to align sequential frames, enabling auto-

matic transfer of manually delineated structures from one frame to all others. Indeed, registration techniques facilitate the segmentation of cardiac structures across the entire cardiac cycle without requiring manual contouring at each time frame, a task that would be prohibitively time-consuming for clinical applications [95, 188].

B.2.1. Validation of Deformable Image Registration

To quantitatively evaluate DIR accuracy, the primary methods employed in reference literature are landmark-based and contour-based metrics that require expert clinicians to manually delineate ground truth contours or regions of interest (anatomical segmentations) on both the reference fixed and the moving image [188].

The output Deformation Vector Field (DVF) of DIR, mapping the point-to-point displacements from the moving to the fixed image, is then applied to the boundaries or volumetric segmentations manually contoured on the moving image to automatically align them with their counterparts in the fixed reference frame. The so-called warped or deformed results obtained through contour propagation on the fixed image are then compared to the ground truths manually delineated previously.

The level of similarity is quantified through standard similarity metrics that include surface-distance based measures, such as Hausdorff distance (Hausdorff Distance (HD)) and mean surface distance (Mean Surface Distance (MSD)), and Dice coefficient (Dice Coefficient (DC)) for binary overlap [95, 113, 188].

Hausdorff Distance

In mathematics, the HD serves as a standard metric to quantify the maximum distance between two subsets of points within a metric space [181]. Given two surfaces or sets of points, S_x and S_y, in a metric space, the Hausdorff distance between these sets is defined as follows:

$$HD = \max \left\{ \sup_{x \in S_x} \inf_{y \in S_y} \text{dist}(x, y), \sup_{y \in S_y} \inf_{x \in S_x} \text{dist}(x, y) \right\} \tag{B.1}$$

where $\text{dist}(x, y)$ represents the distance between point x belonging to S_x and point y belonging to set S_y.

Both the maximum distance from a point in S_x to its nearest point in S_y and the maximum distance from a point in set S_y to its nearest point in S_x are taken into account. The HD is then defined as the largest of these two values, meaning the maximum distance between any point in one set and its nearest point in the other set. The lower the Hausdorff

distance, the more similar the two subsets of points are.

In medical imaging, when applying the Hausdorff distance to compare segmented volumes or regions of interest (ROI) in 3D space, the computation is typically performed between all voxels belonging to the external surfaces of the two ROI under analysis, considering $\text{dist}(x, y)$ as the 3D Euclidean distance between closest points [71].

Mean Surface Distance

The MSD is defined as the average of the minimum surface distances from each point on the segmented contour to the nearest point on the ground truth contour and vice versa. Considering two surfaces S_x and S_y, the MSD is computed as:

$$MSD = \frac{1}{N_x + N_y} \left(\sum_{x \in S_x} \inf_{y \in S_y} \text{dist}(x, y) + \sum_{y \in S_y} \inf_{x \in S_x} \text{dist}(x, y) \right) \tag{B.2}$$

where the sum of N_x and N_y is the total number of points in both contours.

A higher MSD between S_x and S_y indicates the existence of regions of dissimilarity between the two sets. In the context of evaluation of DIR accuracy, reference recommendations exist for MSD and mean distance from agreement (Mean Distance from Agreement (MDA)), which measures the same extent but only for a subset of corresponding points between S_x and S_y. To validate an image registration framework, the standard recommendation by the American Association of Physicists in Medicine (American Association of Physicists in Medicine (AAPM)) Task Group (TG) is to ensure that the MSD or MDA is less than 2 mm [28, 113].

Dice Coefficient

In the domain of image segmentation and registration, the DC serves as the predominant reference metric that is commonly employed as a comparative metric to evaluate the accuracy of different DIR algorithms [28, 113]. More precisely, it is a measure to quantify the similarity between two sets of regions or segmentations in terms of voxel-based overlap.

Considering two sets X and Y, the Dice similarity coefficient is computed as:

$$DC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{B.3}$$

where $|X|$ and $|Y|$ are the total number of voxels in X and Y respectively, while $|X \cap Y|$ is the number of voxels that are common to both X and Y.

The DC ranges from 0 to 1, where 0 indicates no overlap (complete dissimilarity) and 1 indicates perfect agreement between the two sets. Dice coefficients above 0.85 are commonly considered acceptable. However, AAPM-TG recommends a DC greater than 0.9 for proper validation of DIR techniques, with values exceeding 0.95 viewed as highly favorable [28, 113].

B.3. Patient-Specific 4DCT Dataset Acquisition

To conduct a complete reconstruction across an entire cardiac cycle, a preoperative 4DCT image dataset was acquired with ECG gating to obtain 21 frames spanning over the entire R-R interval, from 0 to 100%. The cardiac CT sequence was loaded on 3D Slicer and analyzed to identify the reference frames associated with the two extreme cardiac phases: peak diastolic (95%) and peak systolic (40%).

The 4DCT dataset utilized in this work was acquired with the following specifications:

- Spatial resolution: $0.785 \times 0.785 \times 0.600$ mm
- Temporal resolution: 20 frames per cardiac cycle at 5% intervals
- Acquisition mode: Retrospective ECG gating
- Reference frames: EDF at 95% of R-R interval, peak systolic at 40% of R-R interval

The high spatial and temporal resolution of the dataset enabled accurate 3D reconstruction of cardiac structures throughout the cardiac cycle through the deformable image registration technique described below.

B.4. Implementation of Deformable Image Registration for Pericardial Segmentation

In the present work, segmentations for frames beyond the EDF were obtained through deformable image registration, by registering each frame to the EDF used as the fixed reference. The DIR technique was implemented through SlicerElastix Extension Module based on the elastix Toolkit [103]. The primary objective was to propagate the manually segmented pericardium from the EDF to all other frames of the cardiac cycle, enabling complete 3D reconstruction of the pericardial surface across all time points.

The transform ${}^{mov}T_{fix}$ mapping the spatial correspondence between voxels in each moving frame and the fixed diastolic frame was obtained as output. This transform was

then applied to the pericardium segmentation mask manually delineated on the EDF, automatically generating the pericardial segmentation for each frame through contour propagation.

B.4.1. Validation of the Registration Framework

To validate the accuracy of the DIR algorithm and assess the reliability of the registered displacement field, a conventional approach based on comparing registered segmentations with ground truth segmentations was adopted. The output transform ${}^{mov}T_{fix}$ was applied to the manual segmentation of diastolic pericardium to obtain the automatic warped segmentation at peak systolic frame through DVF-contour propagation. The output systolic segmentation was then compared to the ground truth segmentation of pericardium manually contoured on the reference systolic volume.

Typical metrics were computed to measure the extent of similarity in terms of surface distances and voxel-based overlap: maximum and 95th percentile of HD, MSD, and DC. The comparison resulted in similarity metrics that were high above the recommended ranges for DIR validation (DC $>$ 0.9, MSD $<$ 2 mm), confirming the robustness of the image registration technique adopted.

These validation results demonstrated that the registration framework was sufficiently accurate for automatic segmentation propagation across all frames, enabling reliable 3D reconstruction of the pericardium throughout the cardiac cycle as described in the main text.

List of Figures

2.1 Overview of the CAS workflow. The workflow comprises three main phases: **(A)** Images Acquisition: preoperative medical imaging (CT/MRI) of the patient's anatomy; **(B)**Processing and Planning: segmentation and 3D reconstruction of anatomical structures, followed by surgical planning and target identification; **(C)** Intraoperative Phase, which includes: **(1)** Registration and Tracking establishing spatial correspondence between preoperative images and the patient's intraoperative reference frame; and **(2)** Navigation through a 2D multimodal display that integrates preoperative planning data with intraoperative tool tracking, enabling precise navigation of the patient's anatomy. Adapted and extended from [163]. 9

2.2 Sensor-based tracking system technologies [163]. (A) Optical Tracking System: NDI Vicra (NDI - Northern Digital Inc., ON, Canada) infrared camera tracking a passive marker probe; tip position $p_{tip,OT}$ determined through calibration. (B) Electromagnetic Tracking System: NDI Aurora® field generator tracking an EM sensor probe; tip position $p_{tip,EM}$ directly estimated. (C) Fiber Bragg Grating Shape Sensing, optical fiber (30–50 cm) with strain sensors; wavelength shifts enable real-time shape reconstruction. Reproduced with permission from Authors [163]. 14

2.3 The Medical Extended Reality Framework, as outlined by Navab et al. [157], comprises four core components. The system perceives the physical world through its sensing modalities, constructs a *Digital World* representation by processing the collected information, and presents augmented content through the *XR Display* that users perceive and interact with via *XR User Interaction* interfaces. *Evaluation* is essential throughout the system's design and development. Reproduced with permission from Authors [157], published in Journal of Imaging, ©2022 MDPI. Licensed under an open access Creative Commons CC BY 4.0 license. 21

3.1 Medical XR Framework [157] enhanced with Surgical Interaction Modeling layer. The framework extension introduces a physics-based model of tool-tissue interaction positioned between the Physical World and the traditional Digital World representation. This model processes sensor-derived observations from the Physical World (catheter shape, instrument tracking) and generates two distinct outputs: (1) *Anatomical Updates* that refine the Digital World representation by predicting unobservable tissue states (vessel deformations, temporal cardiac motion), maintaining spatial correspondence under observational constraints; and (2) *Interaction Features* that encode dynamic properties of the surgical interaction (proximity to moving structures, contact forces, tissue resistance) and directly feed both Visual Display and Auditory Display modalities within the AR/VR Display layer. This dual-output architecture addresses both the spatial challenge (predicting what cannot be directly observed) and the perceptual challenge (delivering high-temporal-resolution guidance through complementary sensory channels). The Physical World characteristics particular to cardiovascular PIs, soft tissue deformability, highly dynamic cardiac beating, and limited observability, motivate the need for this physics-based interaction modeling approach. 39

4.1 Schematic overview of the proposed framework. An *Initialization* step preprocesses vessel (A) and catheter (B) models and instantiates the interactive simulation. A *Registration* step aligns sensor ($\{EM\}$) and anatomical ($\{CT\}$) frames (C). During *Runtime*, the interactive simulation continuously updates the XR visualization (D). 51

4.2 Experimental setup for implementing and testing the framework. A robotic driver inserted the sensorized catheter into a silicone phantom replicating the vessel anatomy. Two high-speed cameras tracked surface-marker displacements on the phantom wall, while the corresponding FEM nodes were identified by applying the transformation ${}^{CT}\mathbf{T}_{EM}$ to the point cloud acquired at rest using the EM probe. 52

4.3 A 7.4 mm catheter **(A)** housed a multi-core FBG fiber and two EM sensors within its central lumen **(B)**. The sensor-based reconstructed catheter centerline, expressed in the $\{EM\}$ reference frame as $\mathbf{c}_{\mathrm{EM}}(\underline{x}, t)$, was uniformly discretized into 78 points. Each point served as the center of a rigid sphere, and consecutive points were connected by cylinders of 7.4 mm diameter, producing a piecewise-continuous geometric avatar of the catheter outer surface for collision and contact handling **(C)**. 54

4.4 Mock insertion trajectory for computational benchmarking. A fictitious catheter trajectory generated via NURBS interpolation of the vessel centerline was used to simulate insertion at 10 mm/s. **(A)** The LM contact model maintains stability and correctly computes the deformed vessel configuration induced by the complex trajectory. **(B)** Displacement field magnitude for the baseline configuration ($E = 1.0$ MPa, $\nu = 0.36$), serving as the reference for all sensitivity analyses. The region of interest (ROI) indicates the zone used for localized displacement metrics, while global metrics were computed over the full vessel lumen. 57

4.5 Interactive FEM performance: simulated time (T_{CPU}) vs. physical time (T_{PHYS}) for three repeated tests. Three latency phases were identified: startup, low-latency (slope 1.12), and high-latency (slope 1.45). The deformed vein lumen, color-coded by displacement magnitude, is shown at specific time-points. 63

4.6 **A)** Ground-truth marker configurations ($\mathbf{p}_{EXP}$) overlaid on the undeformed vessel (light gray) at three representative time-points; each is annotated with the 95th percentile of the displacement magnitude ($\|\boldsymbol{u}\|_{95}$). **B)** Corresponding FEM configurations at the same time-points, shown with a displacement colormap and overlaid on the undeformed vessel. The $\|\boldsymbol{u}\|_{95}$ computed over the marker-covered region is reported. **C)** *Relative displacement error* for the three repeated tests, expressed as $\Delta d(\underline{x}_i, t_j)$. Median errors across the 13 time points are shown, with shaded bands indicating the 5th-95th percentile range. Representative FEM vessel-catheter configurations are shown at the three time-points. 64

5.1 Posterior-based (A) and anterior-based (B) approach for epicardial puncture. Needle's insertion under standard fluoroscopic views with pericardial silhouette annotation (C, D). Adapted from [179]. 72

5.2 High-level schematic of the proposed workflow. 76

5.3 Overview of the XR-based multisensory surgical navigation system developed for PEA. (A) ECG-gated 4DCTA frames over one cardiac cycle were analyzed. The myocardium and the external pericardial membrane (i.e., pericardium) were segmented, and their three-dimensional models reconstructed. An *ad hoc* planning algorithm was then applied to compute risk-optimized needle trajectories, avoiding surrounding critical organs. (B) Real-time computation of navigation parameters, defined as the Euclidean distance between the needle tip (x_{tip}) and the moving pericardium (d_{TP}) and myocardium (d_{TM}). A navigation setup for real-time needle tracking was implemented. Registration steps were required to coherently track the needle tip (x_{tip}) relative to the CT-derived phantom model ({CAD}), enabling the computation of navigation parameters. (C) The XR-based multisensory display integrated modular visual and auditory components with input data prepared as detailed in (B), to provide accurate and dynamic tool guidance. 77

5.4 Schematic illustration of the *ad hoc* algorithm's workflow to isolate safe insertion trajectories for patient-specific epicardial punctures. The algorithm begins with user selection of either an anterior or posterior approach, restricting the search to the associated pericardial surface. Three primary segmentations are automatically classified: critical structures to avoid during needle insertion, skin access surface, and target pericardial area (anterior or diaphragmatic portion). An oriented bounding box tree locator is initialized for collision detection. The user places markers on the skin segmentation to define a custom access region. Access and target areas are discretized with a resolution of $\frac{1\,\text{marker}}{0.5\,\text{cm}^2}$, distributing potential puncture and target sites accordingly. For each target site, the algorithm loops through all puncture sites to isolate risk-free trajectories satisfying the three filters. All trajectories are displayed with key measurements: length, angular inclination, and minimum distance from critical organs. 79

5.5 Conceptual representation of anatomical regions in PEA for state-based sonification. Transitions are defined based on the Euclidean distance computed along the needle axis from $\mathbf{x}_{tip}$ to both the moving pericardium (d_{TP}) and myocardium (d_{TM}). The figure is for illustration purposes only; size and order of magnitude are not to scale. 81

5.6 Planned trajectories for the phantom study, shown at End-Diastolic and Peak-Systolic phases of the cardiac cycle. Six trajectories were generated toward three target points (1,2,3) on the pericardium, following the strict requirements of PEA (Section 5.2.1). The myocardium and the initial tract of the pulmonary artery are shown in gray. The pericardium is color-coded to represent the inter-parietal distance between the myocardium and pericardium. This distance varies both spatially and temporally: it increases during Peak Systole, whereas End Diastole represents the most critical phase, with the myocardium at maximum volume and a higher risk of myocardial injury. 88

5.7 Distribution of task outcome rates (*Successful Completion*, *Missed Target*, and *Critical Failure*) under visual-only (V) and MS modalities for all participants, novices, and experts. The stacked bars highlight the higher success rates and lower error rates under MS across all groups. 92

5.8 Box plot visualization of successful needle placement accuracy evaluated under two aspects : (A) stopping advancement upon reaching the pericardial membrane; and (B) accurately reaching the pre-planned target points (). The n variable represents the number of samples considered. 93

5.9 Final successful needle tip placements for both modalities relative to the pericardial membrane. Colors indicate the minimum distance to the dynamically moving pericardium (evaluated across 20 frames). The multisensory condition (bottom) exhibits greater spatial consistency, with placements more closely aligned to the membrane. 94

5.10 Needle tip placement relative to the planned target points 1, 2, 3 across both modalities for successful placements only. The value of considered samples for each test is reported (n). **(A)** 3D distribution of final tip positions relative to the pericardial surface (End-Diastolic phase). The clustering effect demonstrated by the MS modality, helps to reduce high-risk placements with tips avoiding critical areas of cardiac wall proximity. **(B)** Each subplot shows the spatial spread of final tips. While both modalities exhibit deviations from the planned targets, MS generally resulted in tighter clustering and reduced spread compared to V. 95

5.11 Modality-dependent time-accuracy relationships. Needle placement accuracy is measured by the distance of the needle tip position to the planned target point. The less distance represents better performance. Visual-only trials (gray) exhibit a positive correlation ($\rho = +0.31$), while multisensory trials (blue) show a negative trend ($\rho = -0.28$). 96

6.1 BioSonix Design Framework – A General Overview 106
6.2 FEM-based simulations model tool-tissue dynamics by simulating insertion trajectories in a simplified setup. (A) The *Anatomical Domain* includes skin, fat, and muscle, discretized into tetrahedral elements. (B) The tool insertion is defined by an entry point, orientation, and a velocity $\vec{\boldsymbol{v}}$ applied to the handle. The $\boldsymbol{p}$-axis, aligned with the tool, spans the model, where nodal displacements $\mathbf{u}(\mathbf{x}_{A_i}, t)$ are computed along specific nodes (light blue dots) and mapped to the *Sonification Domain* via M, T, and I. (C) A 1D spring along the $\boldsymbol{p}$-axis represents the basic sound model, while (D) a more advanced model expands orthogonally, forming a 3D topology centered on the $\boldsymbol{p}$-axis. 110
6.3 Examples from the systematic experimentation to optimize the sound model. (A, B, C) Results from the *Nodal Characterization* experiment, demonstrating the correlation between displacements in the *Anatomical Domain* and the resulting sound. (D, E) Results from the *Sonic Area Definition* experiment, illustrating the effects of varying the distance between *Drivers*. (F) A sample from the *Nodal Contribution Impact* experiment, showing the gradient achieved by activating multiple nodes. 113
6.4 Illustrative examples of the BioSonix design framework applied to abstract models. (A, B) Demonstrate the framework's stability and responsiveness when tissue classes and their deformability vary. Spectrograms reveal transitions between tissues, with signal intensity reflecting tissue stiffness. (C) Highlights the framework's ability to differentiate dynamic tool-tissue interactions through sound signals, simulated via varied tool trajectories. . . 114

List of Tables

4.1 Mesh sensitivity analysis results. T_{CPU} is total accumulated machine time, and $|\mathbf{U}_{95}|$ is the 95^{th} percentile of displacement magnitude at configuration $\Gamma_{\bar{t}}$. 58

4.2 Cumulative physical time (T_{PHYS}), simulated time (T_{CPU}), and lag at sampled time-points. Ranges indicate variability across three experimental repetitions. 62

5.1 State-Dependent Parameter Modulation. The normalized input variables $\hat{d}_{TP}$ and $\hat{d}_{TM}$ define the interpolation intervals within each state. Membrane pitch is reported as fundamental frequency f_0, which operationalizes membrane radius R. Temporal excitation intervals Δt are given in milliseconds. 84

5.2 Task outcome rates for expertise group under visual-only (V) and MS modalities. All the enrolled participants (12) performed 5 tasks per modality. Experts (5) and novices (7) subgroups were also analyzed. The higher values by *Success* and the lower values by *Missed Target* and *Critical Failure*, which show better performance, are bolded. 91

5.3 Summary of accuracy measurements for final needle tip placement. Values are reported for successful trials only. n is the number of samples considered for each modality. 92

6.1 Mechanical properties of various tissue classes, including the elastic modulus (E) in megapascals (MPa) and density (ρ) in kilograms per cubic meter (kg/m^3). 116

6.2 The statistical results of the stopping task across all cases are presented, including Exact Match (%), the proportion of exact matches as a percentage. High-Performing values are bolded based on thresholds: Mean values close to 0 ($-0.5 \leq$ Mean $\leq +0.5$), Std. Deviation below 1.0, and Exact Match (%) above 70%. Negative values indicate that stopping occurred before the target, positive values indicate stopping after the target (in seconds), and zero indicates stopping precisely at the target. The notation T# refers to the trajectory number, while the second value indicates the thickness of the critical tissue layer (in millimeters). 117

A.1 Mechanical parameters sensitivity analysis: variation of Young's modulus (E) and Poisson's ratio (ν) from baseline configuration ($E = 1.0$ MPa, $\nu = 0.36$). Metrics quantify local displacement deviation ($\Delta\mathbf{U}$) and global shape change (Chamfer Distance, *ChDist*). 153

Acknowledgements

This doctoral thesis was supported by MUSA - Multilayered Urban Sustainability Action - project, funded by the European Union - NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.5: Strenghtening of research structures and creation of R&D "innovation ecosystems", set up of "territorial leaders in R&D".

List of Acronyms

2D two-dimensional

3D three-dimensional

4DCT 4-Dimensional Computed Tomography

4DCTA 4D Computed Tomography Angiography

AAPM American Association of Physicists in Medicine

AI Artificial Intelligence

AR Augmented Reality

BC Boundary Condition

CAS Computer-Assisted Surgery

CBCT Cone Beam Computed Tomography

CT Computed Tomography

CTA Computed Tomography Angiography

DC Dice Coefficient

DIR Deformable Image Registration

DOF Degrees of Freedom

DVF Deformation Vector Field

ECG Electrocardiogram

EDF End-Diastolic Frame

EM Electro-magnetic

FBG Fiber Bragg Grating

FDA Food and Drug Administration

FEM Finite Element Model

FoR Frame of Reference

FOSS Fiber Optic Shape Sensing

FPS Frames per Second

H2 Hololens 2™

HD Hausdorff Distance

HMD Head-mounted Display

ICE Intracardiac Echocardiography

IQR Interquartile Range

IVUS Intravascular Ultrasound

LM Lagrange Multiplier

MAD Mean Absolute Deviation

MBS Model-based Sonification

MDA Mean Distance from Agreement

MR Mixed Reality

MRI Magnetic Resonance Imaging

MS Multisensory

MSD Mean Surface Distance

MSM Mass Spring Model

NASA-TLX NASA Task Load Index

NCA Neural Cellular Automata

NURBS Non-Uniform Rational B-Spline

OR Operating Room

OSC Open Sound Control

OST-HMD Optical See-Through Head-mounted Display

OTS Optical Tracking System

PEA Percutaneous Epicardial Access

PI Percutaneous Intervention

PLAAC Percutaneous Left Atrial Appendage Closure

PMSon Parameter-mapping Sonification

RMSE Root Mean Square Error

ROI Region of Interest

ROS Robotic Operating System

SHD Structural Heart Disease

SLAM Simultaneous Localization and Mapping

SOFA Simulation Open Framework Architecture

SVD Singular Value Decomposition

TAVR Transcatheter Aortic Valve Replacement

TEE Transesophageal Echocardiography

TEER Transcatheter Edge-to-edge Repair

TTE Transthoracic Echocardiography

VR Virtual Reality

VST-HMD Video See-Through Head-mounted Display

XR Extended Reality