ROWSIM: DESIGNING AND EVALUATING AMBIENT INTERACTION IN MIXED-REALITY ROWING

by

Arsh Shah

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia

August 2026

I begin in the name of Allah, the Most Beneficent and the Most Merciful.

I dedicate this work to Allah, Al-ʻAlīm, the All-Knowing,
and Al-Ḥakīm, the All-Wise.

To the One before whom every question begins and every answer remains incomplete.

We spend our lives drawing the unknown into the known—observing, reasoning, building, naming. Yet knowledge does not diminish what is unknown; it reveals how much more there is to understand.

May I never confuse understanding with mastery, certainty with truth, or completion with an end.

And may every answer leave me with the humility to ask a better question.

رَبِّ زِدْنِي عِلْمًا

"My Lord, increase me in knowledge."

— The Qur'an, 20:114

Alhamdulillah.

Contents

List of Tables vii

List of Figures ix

Abstract x

List of Abbreviations Used xi

Acknowledgements xii

1 Introduction 1

1.1 Context and Motivation 1
1.2 Research Problem and Objectives 2
1.3 Contributions 3
1.4 Thesis Structure 4

2 Literature Review 5

2.1 An Anomaly at the Centre of Indoor Rowing 5
2.2 Designing for the Periphery 6
2.2.1 Calm technology and the traffic between centre and periphery . . . 6
2.2.2 Foreground and background as a design axis 8
2.2.3 Architectural surfaces as a display medium 9
2.2.4 Ambient data visualization, and displays that remember 10
2.3 Against Prescription: Situated Action and Ambiguity 11
2.3.1 Plans as resources rather than determinants 11
2.3.2 Ambiguity as a design resource 12
2.4 Feedback, Skill, and the Meaning of Variability 13
2.5 Engagement as a Measurable Construct 15
2.6 Ambient and Mixed Reality Feedback for Exercise 16
2.7 Ergometer Measurement and Rowing-Adjacent Systems 16

2.8 The Gap This Thesis Addresses 18

3 System Implementation **20**

3.1 Design Goals and Chapter Scope 20
3.2 Design Rationale 20
3.3 Physical Installation 24
3.4 System Architecture 27
3.5 Runtime Execution Model 29
3.6 Fluid Simulation Model (Summary) 30
3.7 Per-Frame Render Graph 32
3.8 Sensor Architecture and Event Normalization 33
3.9 Calibration, Stroke Ownership, and Rowing Physics 34
3.10 Visualization and Ambient Feedback Mapping 40
3.11 Performance Engineering and Empirical Validation 44
3.12 Reproducibility, Assumptions, and Limitations 45
3.13 Summary 46

4 Methodology **47**

4.1 Research Design 47
4.2 Participants 49
4.3 Materials and Setting 52
4.4 Instruments 54
4.5 Procedure 56
4.6 Measures and Derivation 57
4.7 Analysis Plan 58
4.8 Qualitative Analysis 61

5 Results **63**

5.1 Participants and Sessions 63
5.2 Engagement and Experience 63
5.2.1 Questionnaire outcomes 63
5.2.2 Absorption is not the same as losing oneself 64
5.3 Rowing Behaviour Under an Open Task 66
5.4 Casual and Elite Rowers Compared 67
5.4.1 Where the groups did differ 70
5.5 Associations Between Experience and Behaviour 72
5.6 Qualitative Findings 76

5.7 Integration of the Three Strands . . . 83
5.8 Colour and the Refusal of Realism . . . 86
5.9 Comfort and Accessibility . . . 87

6 Discussion **89**
6.1 Interpretation of Findings . . . 89
6.1.1 Engagement did not track a large difference in rowing . . . 89
6.1.2 The display was reported as non-disruptive, and an expert named the mechanism . . . 90
6.1.3 Ambiguity invited exploration, and the mechanism was mostly calibration . . . 91
6.1.4 Displacing effort is not the same as supporting training . . . 93
6.1.5 An expert convergence on the problem with numbers . . . 94
6.1.6 Stopping was not something the rower could do . . . 95
6.1.7 The open task made behaviour interpretable . . . 97
6.1.8 Two engagement modes, and an instrument that cannot see them . . . 98
6.2 Comparison to Literature . . . 98
6.2.1 Does calm computing survive exertion? . . . 98
6.2.2 Ambient displays . . . 99
6.2.3 Mixed reality and exergame research . . . 100
6.2.4 Rowing and sport technology . . . 100
6.3 Limitations . . . 101
6.4 Future Work . . . 104
6.5 Design Considerations . . . 111

7 Conclusion **113**

A Extended System Implementation Detail **132**
A.1 Code and Data Availability . . . 132
A.2 Metal Resource Architecture . . . 133
A.3 Fluid and Wave Solver: Full Discretization . . . 134
A.3.1 Advection . . . 134
A.3.2 Divergence and Pressure Projection . . . 135
A.3.3 Vorticity Confinement . . . 136
A.3.4 Wave / Ripple Field . . . 138
A.3.5 Bloom and Particles . . . 138
A.4 Compute / Fragment Branch Divergence . . . 138

A.5 PM5 Telemetry: Service and Characteristic Layout 139
A.6 Inertial Sensing: Firmware and Signal Layout 140
A.7 Runtime Configuration Surface 142
A.8 Additional Installation and Output Views 143
A.8.1 The caustic layer, and what participants made of it 143
A.9 Session Logging Artefacts 146

B Instruments and Scoring **147**
B.1 Engagement Questionnaire 147
B.2 RowSim Experience Questionnaire (REQ) 148
B.3 Internal Consistency of the Two Scales 150
B.4 Interview Guide 152
B.5 Demographic and Background Questionnaire 152
B.6 Research Ethics Board Approval 154

List of Tables

3.1 Key design and engineering decisions 23
3.2 Subsystem responsibilities 27
3.3 Sensor and physics constants 39
4.1 Participant characteristics by group 52
4.2 Instruments considered and not administered 56
5.1 UES-SF subscale scores 64
5.2 REQ item-level means by group 65
5.3 Casual and elite comparison 68
5.4 Per-participant summary 72
5.5 Associations between engagement and rowing behaviour 73
5.6 Integration of quantitative, behavioural and qualitative evidence 85
A.1 Notation used in the solver discretization 135
A.2 Compute vs. fragment-fallback divergence 139
A.3 PM5 BLE characteristics consumed 139
A.4 libmapper signals consumed from the inertial unit 140
A.5 Inertial stroke-detection parameters 142
A.6 Environment variables recorded per session 143
B.1 Engagement questionnaire as administered, against the published UES-SF . 148
B.2 REQ items as administered 150
B.3 Internal consistency of the two scales as administered 151
B.4 Demographic and background questionnaire fields 153

List of Figures

1.1 Thesis roadmap 4

2.1 Conceptual feedback loop motivating the study 6
2.2 The four traditions this chapter assembles 6
2.3 Design space for rowing and ambient feedback 19

3.1 Projection surface trials 25
3.2 Derivation of the projection mask 25
3.3 RowSim physical installation 26
3.4 End-to-end RowSim architecture 28
3.5 Platform split across two hosts 29
3.6 Runtime execution and concurrency model 30
3.7 RowSim per-frame render graph 33
3.8 Sensor-fusion and interaction pipeline 34
3.9 Handle-mounted IMU and its enclosure 35
3.10 What a stroke forks into 36
3.11 Stroke ownership and arbitration 37
3.12 Visualization composition pipeline 41
3.13 RowSim ambient visualization output and palette variants 42
3.14 Ingest latency relative to one frame interval 45

4.1 Ergometer and console cover 53
4.2 Participant's-eye view of the ambient projection 54
4.3 Session flow for the RowSim user study 57
4.4 Mapping research questions to instruments and analyses 59

5.1 Session timelines for all seventeen participants 66
5.2 Stroke rate against work per stroke, by group 69
5.3 Group separation across every measure tested 70
5.4 Physical output and engagement, by group 71
5.5 Item-level response distributions for both instruments 75
5.6 Interview themes and the design commitments they bear on 76

6.1 Physiological input under development 105
A.1 GPU frame capture confirming slab format and resolution 134
A.2 libmapper signal graph as connected during a session 141
A.3 Image-background mode and tilt-dependent caustics 145
A.4 The blue-on-black configuration from four viewpoints 145
A.5 Session logging bundle structure 146
B.1 Research Ethics Board letter of approval 154

Abstract

A rowing ergometer reproduces the effort of locomotion while removing its perceptual accompaniment, filling the gap with a console that must be read. Ambient-display research has argued since Xerox PARC that information can occupy the periphery, but almost entirely for people at rest.

RowSim augments an ergometer with a floor-projected fluid simulation driven by the rower's movement. Effort scales the wake's brightness and how hard it pushes the fluid; nothing is a number or recoverable as a value. Seventeen rowers, twelve casual and five elite, rowed an open-ended ten-minute session with the console covered, then answered questionnaires and an interview.

The groups differed by a factor of 3.7 in mean power, with all seven physical measures separating them; no engagement measure did. Two participants independently reported that the flow did not read as their own, because no action could stop it: restoring optic flow is a coupling problem, not a rendering one.

The contribution is a working system, a protocol whose refusal to prescribe turns behaviour into evidence, and evidence that a peripheral, non-symbolic display survives exertion.

List of Abbreviations Used

Abbreviation	Meaning
AR	Augmented Reality
BLE	Bluetooth Low Energy
CFD	Computational Fluid Dynamics
CFL	Courant–Friedrichs–Lewy (numerical stability condition)
CPU	Central Processing Unit
FGS	Faculty of Graduate Studies (Dalhousie University)
FPS	Frames Per Second
GEM Lab	Graphics and Experiential Media Lab (Dalhousie University)
GPU	Graphics Processing Unit
HCI	Human–Computer Interaction
IMI	Intrinsic Motivation Inventory
IMU	Inertial Measurement Unit
MR	Mixed Reality
MSAA	Multisample Anti-Aliasing
NASA-TLX	NASA Task Load Index
PM5	Concept2 Performance Monitor 5 (ergometer console)
RBGS	Red–Black Gauss–Seidel (pressure-solver method)
REB	Research Ethics Board
REQ	RowSim Experience Questionnaire (custom instrument)
RQ	Research Question
RtD	Research through Design
SAR	Spatial Augmented Reality
SART	Sustained Attention to Response Task
SDT	Self-Determination Theory
SPM	Strokes Per Minute
SUS	System Usability Scale
TCPS2	Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2nd ed.
UDP	User Datagram Protocol
UES-SF	User Engagement Scale – Short Form
VR	Virtual Reality
WXGA	Wide Extended Graphics Array

Acknowledgements

A thesis has a final version, but it rarely has a moment when it becomes finished.

This one came together through many iterations—of the software, the experiment, the writing, and, perhaps most importantly, my understanding of what I was actually trying to build. What began as an idea about rowing and interaction became a larger exercise in asking what technology can mean when it becomes part of an experience rather than simply a tool within it.

I am deeply grateful to the people who were part of that process.

My deepest thanks go to **Dr. Joseph Malloch**, my supervisor, for giving me the space and freedom to pursue an idea somewhere between engineering, design, and research. I am especially grateful for his patience in hearing me work through questions and uncertainty, and for the perspective he brought to the process: helping me see a problem differently, question it, and try to make it better. His guidance repeatedly returned me to an important distinction: making something work and understanding why it matters are different pursuits. That distinction shaped not only this thesis, but how I think about research.

Alongside him, I would like to thank **Dr. Derek Reilly** for his guidance, feedback, and critical perspective throughout the project. His questions often arrived at exactly the point where I had become too close to the work to see it clearly, pushing me to step back from the implementation and consider what the work was actually asking. His perspective helped me look beyond whether something functioned toward what it meant, what it contributed, and whether it could be better. I am grateful for those questions, particularly when the technical details threatened to become more interesting than the questions they were supposed to serve. I am equally grateful to the **GEM Lab** and broader Dalhousie community for the discussions, technical help, and shared enthusiasm that made this project feel like more than an individual undertaking. Research is rarely solitary, even when much of it happens alone at a computer.

Participants gave RowSim what no simulation can give itself: human experience. They gave their time to something still unfinished, met its imperfections with patience, and transformed an implementation into research. I am also grateful to everyone who tolerated

the less glamorous parts of the project—the debugging, failed builds, uncooperative hardware, experiments that raised more questions than they answered, and the endless cycle of changing, testing, documenting, and changing again. Research has a way of making small problems surprisingly memorable.

Shahkul, thank you for your patience, encouragement, and unwavering support; for understanding the long hours and the conversations that somehow became discussions about shaders or fluid simulations; and for being there through the uncertainty, frustration, and small victories. Whatever this thesis represents academically, it would mean considerably less without you behind it.

Finally, I am grateful for the opportunity to have spent this time making something that I genuinely cared about. RowSim began as an idea about rowing and interaction and became a much larger lesson in how technical systems, design decisions, research questions, and human experience intersect. I am proud of where it ended up, but even more grateful for everything I learned getting there.

To everyone who helped me get here: thank you.

Chapter 1

Introduction

1.1 Context and Motivation

Indoor rowing is a strange invention. A rowing ergometer reproduces the metabolic cost and much of the biomechanics of locomotion while removing every perceptual sign that locomotion is occurring: the rower works hard, and nothing goes past. Where self-motion is ordinarily specified by the continuous, lawful transformation of the light arriving at a moving observer (48), the ergometer couples strenuous effort to an optic array that does not change. Into that gap the machine inserts a small console reporting split time, stroke rate, distance and power as digits (15, 32). The instrument is precise and genuinely useful, and it is also *symbolic*: it encodes the activity rather than presenting it, and an encoding has to be read, which means it has to be looked at.

Human–Computer Interaction has a long-standing alternative to that arrangement. Writing at Xerox PARC, Weiser and Brown argued that information can occupy the periphery — "what we are attuned to without attending to explicitly" — and objected to screen displays specifically on the grounds that symbols "do not peripheralize well" (128). Buxton made the distinction operational, characterizing foreground channels as high-bandwidth but bursty and background channels as low-bandwidth but persistent (25), and the tangible-media tradition demonstrated that the surfaces of a room can carry the latter kind (64, 130). Almost all of that work, however, was developed for people who were sitting still. Section 2.1 develops this argument in full.

This thesis therefore asks what belongs in the perceptual space the ergometer empties, and explores whether a mixed reality, ambient visualization can occupy it. *Mixed reality* is used here in Milgram and Kishino's sense of a continuum between the wholly physical and the wholly virtual (81): RowSim registers computer-generated imagery to the rower's physical surroundings, spatially and temporally coupled to their movement, and so sits toward the physical end of that continuum. It is not a head-mounted system, and Section 3.2 gives the reasons for that choice; the projection-based branch of the field is the spatial augmented reality tradition (13, 95). Rather than presenting additional numbers or explicit

coaching instructions, RowSim projects an abstract, fluid-like visualization onto the floor in the rower's lower periphery, responding in real time to their rhythm and movement. The intent is to study how such feedback shapes the experience of rowing — engagement, focus, perceived usability — and whether it is associated with measurable differences in performance-related behaviour.

1.2 Research Problem and Objectives

The core problem addressed in this work is understanding how ambient mixed reality feedback relates to both subjective experience and observable performance behaviour in a physically demanding, rhythm-based activity.

This thesis investigates the following research questions:

- **RQ1:** How do rowers experience and engage with a mixed reality rowing simulation that provides ambient feedback?
- **RQ2:** How does rowing performance during RowSim use differ between casual rowers and elite rowers?

RQ1 is phrased in terms of how the system is experienced rather than what it causes: participants rowed with RowSim active and with the ergometer's own numeric console covered (Section 4.3), so the study characterizes the experience of the ambient display rather than isolating it experimentally against an alternative. RQ2 asks how rowing itself, and its relationship to reported experience, differs between the two groups.

The rowing task was deliberately left open: no performance target, no duration to sustain, no instruction to row continuously, and an explicit invitation to stop whenever they wished (Section 4.1). This matches the protocol to the artefact — a study of a display that declines to prescribe should not itself prescribe — and it means that everything logged is a choice rather than compliance, which makes how long participants rowed and whether they stopped into behavioural evidence in their own right.

The study objectives are to (i) design and deploy a robust mixed reality rowing system suitable for in-person evaluation, (ii) measure user experience and engagement using a contextually adapted form of the User Engagement Scale–Short Form together with a purpose-built experience questionnaire, (iii) collect objective performance and interaction data during a RowSim session, and (iv) compare outcomes between casual and elite rowing populations.

1.3 Contributions

This thesis makes the following contributions.

- The design and implementation of RowSim, a mixed reality rowing system combining an ergometer sensing setup with a floor-projected ambient fluid simulation that responds in real time to rowing activity (Chapter 3).
- A mixed-methods study protocol for evaluating ambient feedback in embodied interaction under an open-ended task, in which the absence of a prescribed target turns rowing duration and voluntary pausing into behavioural measures alongside self-report (Chapter 4).
- An empirical evaluation with seventeen rowers, twelve casual and five elite, in which no detectable difference in engagement with the ambient display accompanied a group difference of nearly fourfold in power output, together with five design considerations for ambient sports interfaces (Chapters 5–6).
- A perceptual account of why such a display is needed at all, and one finding that tests it. The ergometer is treated here not as a machine lacking a screen but as a perceptual anomaly — sustained locomotor effort with an optic array that does not change — and the display as an attempt to restore the missing half. Two participants at opposite ends of the expertise range independently reported that the restored flow did not read as their own, because it could not be stopped by any action. That observation is predicted by the framing, is traceable to a specific property of the physics controller, and identifies deceleration as an unmet condition of perceptual agency in ambient exercise displays (Section 6.1.6).

Taken together, these are not four independent deliverables but one account of how ambient interaction can be operationalized for a physically demanding, rhythm-based activity: a sensor-fusion and stroke-arbitration architecture that keeps performance data trustworthy (Section 3.9), a mapping from rowing biomechanics to an intentionally abstract visual response (Section 2.3), and a mixed-methods protocol suited to comparing how that response is experienced across expertise levels.

The underlying real-time fluid solver follows established stable-fluids techniques (43, 113, 116); it is not itself a novel numerical method. The technical contribution of Chapter 3 is the surrounding system: sensor fusion across heterogeneous telemetry sources, an explicit stroke-ownership arbitration model, and a mapping from rowing biomechanics to an ambient, intentionally abstract visual response.

1.4 Thesis Structure

Chapter 2 reviews related work in ambient and peripheral feedback, mixed reality exercise systems, and rowing ergometer performance measurement. Chapter 3 presents the RowSim system architecture and implementation. Chapter 4 details the study design, participant recruitment and eligibility, data collection procedures, and analysis plan. Chapter 5 presents quantitative and qualitative results. Chapter 6 discusses implications for the design and evaluation of ambient feedback in physical activities, together with limitations and future work. Chapter 7 concludes the thesis. Figure 1.1 shows how the chapters depend on one another and where each research question is set up, operationalized, and answered.

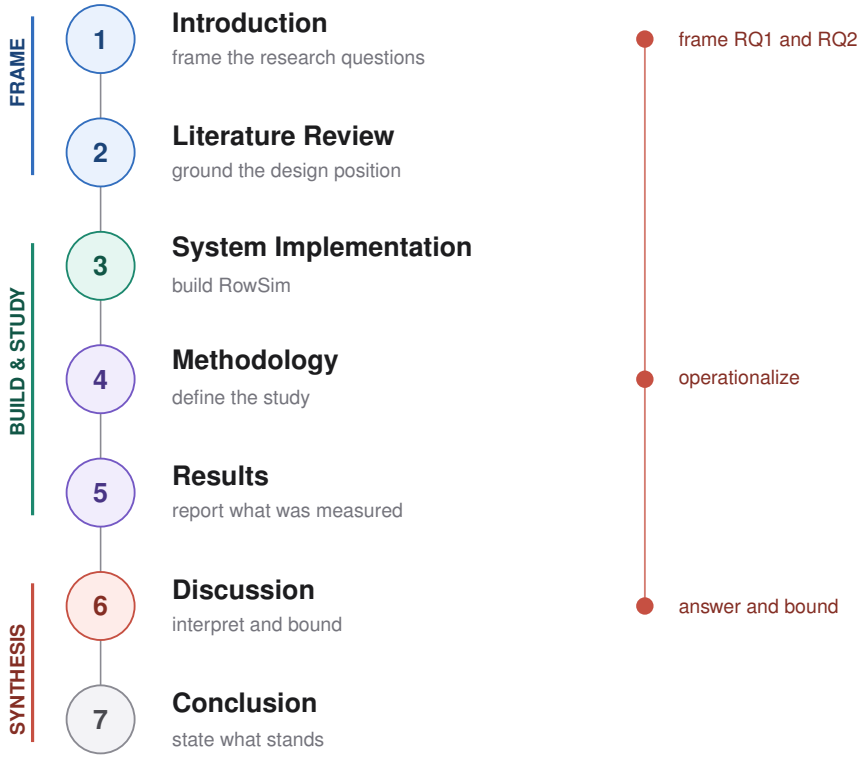


Figure 1.1: Thesis roadmap. The spine gives the reading sequence and each chapter's role in one line; the rules at left group the argument into three phases, and the track at right marks where the two research questions are framed, operationalized and answered.

Chapter 2

Literature Review

2.1 An Anomaly at the Centre of Indoor Rowing

Chapter 1 described the ergometer as a machine that keeps the effort of locomotion and discards its perceptual accompaniment. That description needs one more step before it can do any work, because it is not obvious why the missing accompaniment should matter to a person whose goal is exercise rather than travel.

It matters because self-motion is not ordinarily something a person infers. On Gibson's ecological account, the information specifying it is available directly in the structure of the ambient optic array — the light converging on a point of observation, which for a moving observer undergoes lawful, continuous transformation (48). Work in that tradition has since shown that people use such flow to steer and to hold themselves upright in real time, and that perturbing it disturbs both (10, 70, 126). Optic flow is not decoration on locomotion; it is part of how locomotion is controlled. Rowing hard on an ergometer therefore puts a perceiver in a position their perceptual system has no ordinary precedent for: every proprioceptive and metabolic sign of vigorous locomotion, and an optic array that does not change. Sustained rhythmic locomotion is, on the anatomical evidence, something human physiology may well have been shaped by (17); what it was never shaped for is doing it while nothing goes past.

The console does not repair this, because it answers a different question. Split time, stroke rate, distance and power are *symbolic*: they encode the activity rather than presenting it, and an encoding must be read. So the design question is not how the console might be improved but what belongs in the space it does not fill. Figure 2.1 shows why the space is worth filling at all. Perception and action form a loop, so a display placed inside it does not merely report on the rowing — it becomes one of the conditions the rowing is produced under. The rest of this chapter assembles an answer from four traditions, each supplying a different part of it and all sharing one assumption that this thesis sets out to test (Figure 2.2).

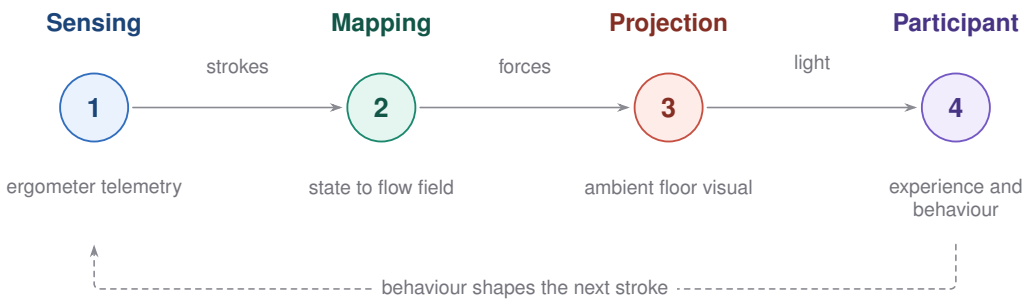


Figure 2.1: Conceptual feedback loop motivating the RowSim evaluation. The display is both a consequence of rowing and an input to subsequent behaviour; this pathway is the design premise examined by the study, not a study result.

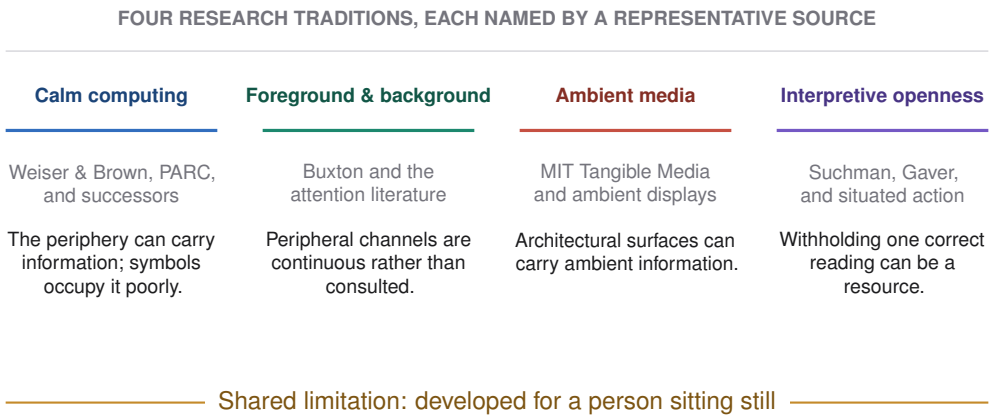


Figure 2.2: The four research traditions this chapter draws on. Each column names a representative source rather than being summarized by it: the traditions are large, and Sections 2.2 and 2.3 work through them in detail. What the four share is not a conclusion but a setting — each was developed for a person who is sitting still. RowSim keeps the peripheral channel and changes the competing demand to sustained physical exertion.

2.2 Designing for the Periphery

2.2.1 Calm technology and the traffic between centre and periphery

That information might live somewhere other than the focus of attention is the founding commitment of ubiquitous computing. Weiser opened his 1991 statement of the programme with a claim about disappearance: “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from

it" (127). Writing at Xerox PARC, he intended this less as a forecast about miniaturization than as a claim about attention — that a mature technology is one that no longer has to be attended to in order to be used.

Weiser and Brown developed the attentional half of that claim into a design position (128). Their definition is precise, and worth stating exactly because later usage often loosens it: "We use 'periphery' to name what we are attuned to without attending to explicitly." So defined, the periphery is a full channel with a different tenancy. "What we mean by the periphery," they insist, "is anything but on the fringe or unimportant." It is the ordinary condition of being informed about one's surroundings — their example is the driver who is not listening to the engine but notices instantly when it sounds wrong. The information was continuously available and continuously used; what it did not require was a decision to consult it.

Two things in that paper bear directly on RowSim, and neither survives citing it merely for the general idea of ambient display.

The first is that calm is not the same as backgrounded. "Calm technology," they write, "engages both the center and the periphery of our attention, and in fact moves back and forth between the two," and the design goal is a technology that will "move easily from the periphery of our attention, to the center, and back." What matters is the traffic, not the residence. A display that could never be brought to the centre would not be calm — it would simply be part of the room, present but unavailable, and a rower who wanted to know what it was doing would have no way to ask. They are equally clear about who should control that traffic: "The individual, not the environment, must be in charge of moving things from center to periphery and back." This distinction governs how Chapter 5 reports participants' accounts of attention. The question is not whether the projection was ignored, but whether attending to it was cheap and leaving it again was easy.

The second is a specific criticism of symbolic displays, which Weiser and Brown make while contrasting their Dangling String — a length of plastic cord twitched by a motor wired into an Ethernet cable, so that network traffic became visible and audible as motion — with the conventional alternative. "While screen displays of traffic are common," they observe, "their symbols require interpretation and attention, and do not peripheralize well." The string, being physically present in the room, has in their phrase "a better impedance match with our brain's peripheral nerve centers." This is as close as the founding literature comes to an argument against the ergometer console. The objection is not that digits are uninformative; it is that digits are structurally ill-suited to peripheral uptake, because reading is not something one does inattentively.

The paper also makes a caveat against itself, which must be reported alongside the rest. "Not all technology need be calm," Weiser and Brown write; "a calm videogame would get little use; the point is to be excited" (128). Rowing at training intensity is not a calm activity, and the corridors and quiet offices in which calm computing was conceived are not physiologically comparable to an ergometer at twenty-five strokes per minute. Whether principles derived from the periphery of a settled attention transfer to the periphery of a body under load is genuinely open. It is, in fact, close to the question this thesis asks.

2.2.2 Foreground and background as a design axis

Buxton made the same distinction operational, in a form precise enough to compare two displays against each other (25). His model separates *foreground* activity — "activities which are in the fore of human consciousness," intentional acts such as speaking or typing — from *background* activity, "tasks that take place in the periphery, 'behind' those in the foreground." His diagnosis of interface complexity follows directly: "a significant amount of the complexity in humans dealing with technology is due to having to explicitly maintain state (or context) as a foreground activity," whereas in ordinary life the maintenance of state is handled peripherally.

The observation that matters most here concerns not attention but the shape of the channel. Buxton notes that foreground activities are characteristically "High Bandwidth but Bursty," while background ones are "Low Bandwidth but Persistent." That contrast describes the two rowing displays exactly. The Concept2 Performance Monitor 5 (PM5), the console fitted to the ergometer used here, is high-bandwidth and bursty: a glance returns four exact quantities, and the glance is short, deliberate, and repeated at intervals. RowSim's projection is low-bandwidth and persistent: what it delivers per instant is small and inexact, but it is delivered without pause and without being asked for. The difference is in the form of the delivery, not the amount — a continuously varying field carries a great deal over a minute, and almost none of it in the form of a value. The two are therefore not competing implementations of one idea. They occupy different regions of Buxton's model, and the claim under test in this thesis is that the persistent low-bandwidth region is under-occupied for exertion — not that it is the better place to be.

Buxton's own emphasis is on movement between the two. "The real power of this model," he argues, "comes not from merely populating the individual quadrants, but by providing the means to make transitions seamlessly from quadrant to quadrant." This study does not test that transition: the console was covered for the duration of the rowing segment (Section 4.3), so participants had the persistent low-bandwidth channel and not the bursty

high-bandwidth one. What it tests is whether the persistent channel is habitable on its own during exertion. Chapter 6 returns to Buxton's transition condition as the design question this study leaves open, and reports an elite participant asking for exactly it.

2.2.3 Architectural surfaces as a display medium

If PARC supplied the attentional argument, the Tangible Media Group at the MIT Media Lab supplied the physical one. Ishii and Ullmer's *Tangible Bits* proposed coupling digital information both to graspable objects at the centre of attention and to the surfaces of the room at its edges, using ambient media — light, sound, airflow, water movement — as background interfaces, and named as its goal the bridging of the foreground and background of human activity (64). Their *ambientROOM* prototype made the walls and ceiling of an enclosure into a low-bandwidth display. Wisneski and colleagues generalized the position: architectural space itself, rather than the screen, becomes the interface between people and digital information (130), an argument advanced in the same period by the Cooperative Buildings programme, which treated the building as the unit of interaction design (117). Dahley, Wisneski and Ishii took the most literal step in RowSim's direction, projecting digital state into architectural space as rippling light — their Water Lamp disturbed a tray of water beneath a lamp so that computational events surfaced as caustics on the ceiling (34).

RowSim belongs to this lineage and differs from it in one respect that structures the whole study. These installations placed continuous, non-symbolic information on the surfaces of calm rooms, for people whose attention was otherwise occupied by office work. RowSim places continuous, non-symbolic information on the floor of a laboratory, for a person whose attention is otherwise occupied by physical exertion. The medium is the same; the attentional economy is not.

The intervening decades supplied evaluation vocabulary for such displays. Mankoff and colleagues adapted heuristic evaluation to ambient displays, discarding the Nielsen heuristics that presume an interactive productivity task and adding ones specific to the domain: sufficient information design, consistent and intuitive mapping, aesthetic and pleasing design, and — directly to the point here — *peripherality of display*. Their recommended set is a hybrid rather than a replacement, retaining about half of Nielsen's alongside the new ones (76); Pousman and Stasko offered a taxonomy along dimensions including information capacity, notification level and aesthetic emphasis (94), and related design work articulated criteria for legibility and non-disruption in everyday settings (1, 90). Bakker, van den Hoven and Eggen subsequently sharpened the underlying construct, characterizing *peripheral interaction* as interaction that can shift fluidly between the centre and periphery of attention

rather than remaining fixed at either (9) — which is Weiser and Brown's traffic condition restated as an empirical property that a system can be assessed against. These criteria were developed for office and domestic environments. Whether they survive translation to a setting where attention is divided between exertion, rhythm and interpretation is one of the things this thesis is positioned to find out.

2.2.4 Ambient data visualization, and displays that remember

There is a specific tradition within this lineage that RowSim belongs to more exactly than to ambient display in general: the use of an aesthetic surface as a data visualization. Redström, Skog and Hallnäs named it *informative art* — works that are "computer augmented, or amplified, works of art that not only are aesthetical objects but also information displays, in as much as they dynamically reflect information about their environment" (96). Their examples map live data onto the visual grammar of existing artworks. The clearest is a Mondrian-style composition of coloured squares, deployed on a screen in the open areas of a university, with about three hundred potential users, showing departures on the local bus line: each square is a bus, its size falls as the departure approaches, and its colour says whether you still have time to walk to the stop (107). Skog and colleagues frame the design problem as a trade-off, and the title states it: between aesthetics and utility. A display that is too legible stops being a thing in the room; one that is too decorative stops carrying data.

RowSim sits on that continuum, and further toward the aesthetic end than informative art usually goes. Where the Mondrian bus display retains a decodable key — this square is that bus, this size is that many minutes — RowSim's mapping is many-to-many by construction (Section 3.10), so no shape, colour or quantity can be looked up. Informative art withholds the *form* of a conventional chart while keeping its decodability; RowSim withholds the decodability too. That is a further step, and it is what makes the ambiguity argument below necessary rather than optional.

One property of the mapping distinguishes it from the informative-art examples more sharply than abstraction does, and it follows from the physics rather than from a display decision. A fluid field is a system with state. Vorticity shed at the catch persists, drifts, spreads and decays over several seconds, so the surface at any instant is not a read-out of the current stroke but an accumulation of recent ones. The display therefore has a short memory, and it holds that memory in a form the rower can see decaying.

Rowing makes this legible in a way few activities would. A rower faces backwards: the direction of travel is behind them, and what is in front of them is where they have already been. A rowing display that puts flow into the space the rower faces is therefore not showing

them what is coming but what they have done — a wake. The strokes visible on the floor at any moment are strokes already completed, receding and dissipating at a rate set by how hard they were pulled. Where a console reports an instantaneous value and then overwrites it, the flow field keeps a few seconds of history in view and lets it go gradually.

That property is doing work in the results. It is the mechanism behind participants' reports of pace and continuity in the absence of any pace read-out, and it is why P10 could attempt a standing start by waiting for the field to run down. Section 5.3 and Section 5.6 return to it.

2.3 Against Prescription: Situated Action and Ambiguity

Deciding that feedback should be peripheral does not determine what it should say. A peripheral display can still be prescriptive — a light that turns red when stroke rate drifts is peripheral and directive at once. The second commitment behind RowSim's design is that the display should not tell the rower what to do. Two literatures argue for that position from different directions.

2.3.1 Plans as resources rather than determinants

Suchman's *Plans and Situated Actions*, written at Xerox PARC, argued that the planning model then dominant in artificial intelligence had the relationship backwards (118). Plans, on her account, are not the mechanism that produces action; they are resources people construct beforehand and invoke afterwards to make action accountable, while the action itself is worked out moment by moment against the circumstances at hand.

Rowing has this character. A coaching prescription — hold the rate, lengthen the drive — is a plan in Suchman's sense, and a useful one. But what a rower does at a given stroke is worked out against fatigue, breath, the previous stroke and the felt state of the body, none of which the prescription anticipates. A display that issues corrections addresses the plan; a display that shows the activity's own dynamics leaves interpretation where the situated account says it already is, with the rower. Dourish's development of this line into a foundation for embodied interaction makes the design consequence explicit: meaning arises in and through engagement with a system rather than being delivered by it, so an interface's job is to afford the making of meaning rather than to transmit a finished one (40). This is why the mapping in Chapter 3 is many-to-many, and why the interview protocol asks what participants took the display to mean rather than whether they read it correctly.

2.3.2 Ambiguity as a design resource

The design position this thesis takes on feedback comes from Gaver, Beaver and Benford's argument that ambiguity, conventionally treated as the enemy of usability, is instead a resource that can encourage close personal engagement with a system (47). Their claim is not that unclear systems are good. It is that in applications outside well-defined work tasks, "traditional concerns for clarity and precision are superseded… by the need to provide rich resources for experience that can be appropriated by users," producing a style of interaction that is "evocative rather than didactic, and mysterious rather than explicit." They distinguish three classes by where the uncertainty sits: *ambiguity of information*, originating in an artefact that does not fully specify what it is showing; *ambiguity of context*, where a thing can be read through more than one frame; and *ambiguity of relationship*, in the individual's own evaluative stance.

The first is the one RowSim uses, and Gaver et al. name the tactic precisely: *use imprecise representations to emphasize uncertainty*. They trace it to Leonardo's instruction that light and shade blend "without lines or borders, in the manner of smoke" — the *sfumato* that softens the Mona Lisa's expression until the viewer must bring it into focus themselves — and extend it to installations that achieve "a kind of digital sfumato… by blurring the usual precision of electronic displays." A rowing console is an unusually precise electronic display. RowSim does not add information to it; it substitutes a representation whose definition is deliberately reduced, a continuous flow field in which no single quantity owns a single visual channel, so that rhythm, continuity and effort remain legible while the exact numeric state does not.

Two consequences shape the rest of this thesis. Ambiguity, Gaver et al. note, "provides a frame of reference that allows the use of inaccurate sensors, inexact mappings, and low-resolution displays because it encourages users to supplement them with their own interpretations" — directly relevant to a system whose fluid behaves plausibly without following hydrodynamic truth (Chapter 3). But they are equally explicit that "ambiguity is not a virtue in itself, nor should it be used as an excuse for poor design. Many ambiguous systems are merely confusing, frustrating, or meaningless." Whether a particular artefact achieves productive under-specification or simple failure is therefore an empirical question, and it is one this study is designed to answer: the questionnaires and interviews are built to capture negative as well as positive responses, and the comparison between casual and elite participants tests whether interpretive openness survives contact with people who already have a technical vocabulary for what they are doing.

One reason under-specification can pay for itself is worth separating from the argument for it, because it concerns what people retain rather than what they enjoy. Design practice has a name for it — *active discovery* — and the claim, as put from the stage at Apple's developer conference, is that "people tend to remember things the best when they experience them themselves, when they learn them through some sort of act of discovery" (3). This is practitioner testimony rather than evidence, and it is cited as such; the experimental literature on generation and testing effects makes a narrower version of the claim under controlled conditions. It is included because it states the design intuition in its own vocabulary, and because it predicts something specific about RowSim that Chapter 6 finds: participants given no account of what the display meant did not wait for one. They constructed readings, disagreed with each other about them, and two of them discovered a property of the system's physics that its author had not articulated.

2.4 Feedback, Skill, and the Meaning of Variability

RowSim's ambient visualization can be understood as a form of augmented feedback: a real-time, interpretable signal linked to user action, presented in an abstract and non-prescriptive form. Evidence across sport and movement contexts suggests that feedback modality and framing influence both performance regulation and motivation. Dobiasch et al. show that alternative visual feedback designs affect an athlete's ability to track target effort profiles during a constant-load cycling task (38), and He and Wei report that augmented reality feedback can improve both motor performance and intrinsic motivation in youth sport training (58). Reviews of feedback for motor learning indicate that its benefit depends on the task, the learner, and the timing of delivery (102, 106), and self-controlled feedback timing specifically has been shown to enhance motor learning relative to externally imposed schedules (65) — a finding that echoes Weiser and Brown's insistence that control of the centre–periphery transition belong to the individual. This builds on a longer tradition: classic schema-theory and knowledge-of-results work established that feedback frequency and guidance shape retention and generalization (102, 104, 129), contextual-interference effects show that practice structure itself alters learning (16, 50), and more recent accounts emphasize attentional focus and learner autonomy as further moderators (133, 134). In SportsHCI, there is growing attention to how "sport data" shapes experience and motivation beyond narrow performance outcomes (93). These perspectives motivate RowSim's dual focus on engagement and experience (RQ1) alongside performance-related behaviour (RQ2).

What stroke-to-stroke variability means

Chapter 5 reports the variability of participants' cycle times, and how that quantity should be interpreted is not self-evident. The default reading treats variation around a target rhythm as error, on the assumption that a skilled performer converges on a fixed movement pattern. Bernstein's work on motor coordination gives grounds for caution (12). Observing that the body has far more mechanical degrees of freedom than any task requires, he argued that skilled action cannot consist in reproducing a stored trajectory, since the same outcome must be achieved from continually changing initial conditions and against continually changing internal states. Practice, on his account, is not repetition of a solution but — in the formulation for which he is best known — *repetition without repetition*: the repeated solving of a motor problem whose terms shift each time it is posed. Variability, on this reading, is partly the signature of a system doing that work, not simply noise added to an ideal.

This does not license treating high variability as a virtue. Rowing is a strongly constrained cyclic task in which stability of rhythm is genuinely part of competence, and the biomechanical literature identifies consistency as one marker distinguishing skill levels (67, 110). The point is narrower and applies to how Chapter 5 argues: a coefficient of variation computed over a single novice session, on a machine most participants had barely used, cannot be read straightforwardly as a measure of technical quality, and this thesis does not read it that way. It is reported as a description of what the rowing looked like, and its association with engagement is examined without assuming which direction a good result would point.

Motivation and optimal experience

A complementary account of why some feedback sustains engagement while other feedback erodes it comes from self-determination theory (SDT), which holds that intrinsically motivated engagement depends on perceived autonomy, competence, and relatedness, and that controlling or heavily evaluative feedback can undermine intrinsic motivation even when it is informative (101). Reviews applying SDT to sport and exercise find that self-determined forms of motivation are associated with persistence, effort, and positive affect, whereas externally regulated or pressured motivation is associated with poorer adherence (115, 122). This bears directly on RowSim's design position, and it sharpens the trade-off already visible in Buxton's terms. A numeric dashboard supplies unambiguous competence information but also functions as a continuous evaluative signal; an abstract ambient display supplies weaker competence information and correspondingly weaker evaluative pressure. Which better supports engagement is an empirical question rather than a settled one, and it is part of what

RQ1 examines. It also motivates the interview protocol's attention to whether participants experienced the projection as responsive to them — supporting competence and autonomy — or as merely decorative.

Sport psychology further highlights experiential constructs such as flow and related optimal-performance states. Csikszentmihalyi's account describes flow as an absorbing, autotelic state arising from a balance of challenge and skill (33); Antonini Philippe et al. build on this with an exploratory qualitative account of how flow emerges, develops, and terminates in elite college athletes and musicians (2). Work on flow and "clutch" states further emphasizes that optimal performance may involve multiple phenomenological modes rather than a single idealized state (119). RowSim does not aim to induce flow directly, but these perspectives motivate measuring attention and absorption, and motivate interpreting differences between elite and casual participants in terms of expertise, expectation, and attentional strategy (77) rather than treating the two groups as interchangeable. The ecological-dynamics tradition in skill acquisition makes the same point from the movement side, treating expertise as an attunement to task constraints rather than as a stored technique (5, 28, 97) — a framing continuous with both Gibson and Bernstein, and one that predicts trained and untrained rowers may pick up different information from the same display.

2.5 Engagement as a Measurable Construct

Because RQ1 asks about engagement, the thesis needs a definition of engagement that can be measured without collapsing into satisfaction or enjoyment. It follows engagement research that treats user engagement as a multidimensional construct with identifiable attributes rather than an informal impression (7, 88). The four attributes used here — focused attention, aesthetic appeal, perceived usability, and reward — are those of the short form specifically: the original scale proposed six, and later factor analyses collapsed novelty, felt involvement and endurability into the reward dimension (89). This sits within broader HCI accounts of user experience as encompassing hedonic, pragmatic, and experiential qualities beyond task success (55, 56, 63, 72). The instrument derived from this work, and the reasons for pairing it with interviews rather than relying on it alone, are described in Chapter 4; the multidimensional structure matters here because it allows the results to distinguish a display that was found beautiful from one that was found absorbing, which turns out to be a distinction the data require.

RowSim's evaluation also sits within a broader ecosystem of rapid feedback systems used in elite training environments (8). That context helps locate the contribution: rather than optimizing explicit coaching signals or dashboards, the system explores how ambient, spatial

feedback might complement existing performance-centric instrumentation by targeting experience and engagement directly, rather than as a by-product of better numbers.

2.6 Ambient and Mixed Reality Feedback for Exercise

Ambient intelligence systems for personalized sport training demonstrate how sensing, real-time data, and decision logic can be combined to shape training experiences (121), and ambient displays have been investigated more broadly as persuasive or behaviour-shaping interventions in ubiquitous computing (99). Within mixed reality — positioned along Milgram and Kishino's reality–virtuality continuum (81) — exergames have been explored for sports rehabilitation, yielding design insights for physically active experiences (51). Interactive floor projection specifically has been used to support physically active experiences in shared spaces (120), narrative-based ambient interfaces have been proposed to reflect and motivate physical activity (86), and wearable ambient displays suggest that low-complexity representations can sustain ongoing engagement (24). Media-facade research makes the complementary point that public displays shape experience and atmosphere beyond pure utility (44).

Exertion games research offers the most developed design perspective on embodied, physically demanding interactive systems, emphasizing bodily engagement, social connectedness, and the balance between explicit goals and open-ended play (78, 82–84). Knaving and colleagues raise a caution particularly relevant here: designing for flow is not by itself sufficient to make an exertion experience meaningful to the people doing it (68). Taken together, these works motivate RowSim's focus on embodied interaction and on feedback that is engaging without being overtly prescriptive.

2.7 Ergometer Measurement and Rowing-Adjacent Systems

Rowing ergometers are commonly evaluated using standardized protocols such as 2000 m time trials and incremental tests. Borges et al. systematically review common ergometer testing protocols and identify the physiological and power-related variables that correlate with 2000 m performance (15). RowSim's study design is not a maximal performance test, but this literature contextualizes which performance-related metrics are meaningful to examine (e.g. timing and rhythm stability) and how such measures might vary with expertise. Biomechanical analyses of rowing technique further identify stroke-level kinematic and

kinetic markers that differentiate skill levels (67, 110, 111), providing a technical vocabulary for interpreting the stroke-level differences between casual and elite participants examined in Chapter 5.

Beyond test protocols, prior work has explored sensing and feedback systems for rowing technique and training. Early work demonstrated a real-time rowing simulator with multimodal feedback (124), and instrumented components such as ultrasonic seat-position tracking have been proposed for technique analysis (49). Reviews synthesize the growing ecosystem of sensor technologies used for rowing biomechanics and performance monitoring (32), and recent wireless measurement systems specifically target ergometer training (60). Most directly relevant to RowSim, Iannucci et al. present ARrow, a real-time augmented reality rowing coach (62). ARrow runs as a tablet application: a coach points an iPad at the rower, computer-vision pose estimation extracts the rower's skeleton from the video, and stroke-phase and technique visualizations are composited onto that live video. The coach is the primary viewer, and a second screen can be streamed to so that the rower sees the feedback while rowing. ARrow is the closest recent system to RowSim in application domain, and the contrast is instructive in exactly the terms this chapter has developed: ARrow delivers explicit, prescriptive coaching on a screen — it marks where the body should be and how far the rower deviates from it — which is a foreground channel carrying a plan in Suchman's sense. RowSim delivers non-prescriptive, continuous information into the lower periphery through floor projection. The two systems address the same activity from opposite corners of the design space.

Research on rowing-adjacent interactive systems has also explored immersive and game-like applications. A single-session study of sixteen rowers reported better performance and experience in a head-mounted virtual-reality condition than without it (6), and virtual reality has also been used to support gamified rowing simulation (73). Rowing has also been used as an interaction technique in multiplayer games (105) and as a modality for combined physical and cognitive challenges (74). In on-water contexts, acoustic feedback has been examined as a tool for perceiving and modifying movement patterns (103) — notable here because sound is a channel that peripheralizes readily, and because five participants in this study raised it unprompted (Chapter 5). A growing body of recent work explores complementary sensing and interaction approaches for rowing specifically: wireless signal-strength optimization for on-boat sensor placement (132), multisource data fusion for indoor training-pool digitization (136), functional data analysis of ergometer motion-capture technique (11), alternative steering and locomotion metaphors for virtual-reality (VR) rowing (59), loss-aversion-driven exergame design (125), coach-facing feedback-solution surveys (112), and head-mounted virtual-reality ergometer platforms with motion tracking for technique training

(123) or on-water rehabilitation (41). While RowSim differs in display modality and in its abstract feedback goals, these projects demonstrate the breadth of feedback and immersion approaches explored for rowing practice and embodied exertion.

More generally, sports data visualization research highlights design challenges in representing heterogeneous performance and tracking data, including tensions between exploratory analysis, communication, and real-time use (91). These concerns are relevant to RowSim even though it does not visualize conventional performance metrics: the system similarly navigates trade-offs between interpretability, attentional demand, and experiential quality.

2.8 The Gap This Thesis Addresses

The literatures above converge on a position that has not been tested where it is hardest. Calm computing established that the periphery is a legitimate place for information and that symbolic displays occupy it badly; Buxton showed that peripheral channels are low-bandwidth but persistent, which is a different kind of channel rather than a worse one; the tangible-media tradition demonstrated that architectural surfaces can carry such channels; Suchman and Gaver argued, from anthropology and from design respectively, that withholding prescription can be a resource rather than a deficiency; and the ecological account explains why continuous flow might be readable where digits are not. Every one of those arguments was developed for a person who was sitting still.

What remains comparatively unknown is how *abstract*, ambient mixed reality feedback is experienced, and how performance-related behaviour unfolds alongside it, during a rhythm-based, full-body activity performed at physiological load — and whether it does so differently for novices and for trained athletes, who arrive with different expectations of what a display owes them.

Figure 2.3 locates the systems reviewed above along the two dimensions that define this gap: how much focal attention a display demands, and how explicitly it encodes performance. Existing rowing feedback clusters in the upper-left region — explicit, and demanding of focal attention — whether it is delivered on a console, in a head-mounted overlay, or in a VR exergame. The lower-right region, where feedback is both peripheral and interpretively open, is comparatively unexplored for physically demanding rhythmic activity. RowSim is a deliberate probe into that region.

This thesis addresses that gap through an in-person, mixed-methods study of RowSim with both casual and elite rowers, combining subjective engagement measures with objective performance and interaction logging.

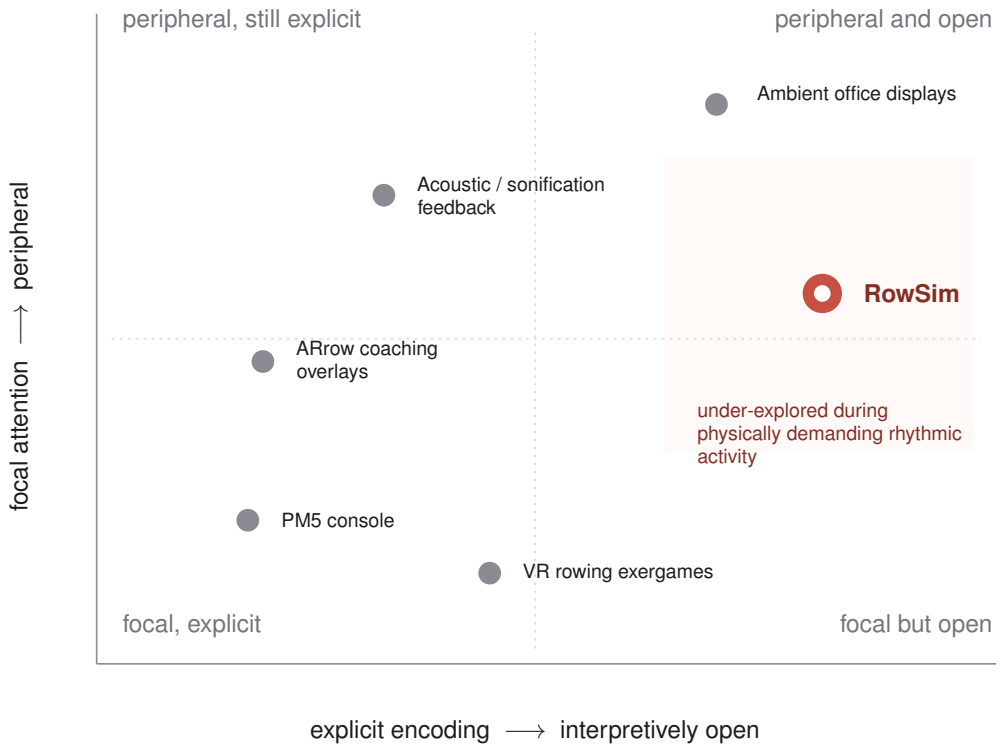


Figure 2.3: Design space for rowing feedback by attentional demand and interpretive openness. RowSim probes the comparatively under-explored combination of peripheral, open feedback and physically demanding rhythmic activity. Positions are qualitative, not measured coordinates.

Chapter 3

System Implementation

3.1 Design Goals and Chapter Scope

RowSim is implemented as a research instrument for studying ambient mixed reality feedback during indoor rowing. Its design is therefore shaped by requirements broader than visual-effect generation alone: the system must sense rowing activity, arbitrate between multiple telemetry sources, map physical movement into a stable interaction vocabulary, execute a real-time fluid-and-wave simulation, and present the result as a floor-projected visualization that remains legible during exertion. These requirements connect the implementation to the thesis's HCI motivation: RowSim is intended to support continuous peripheral awareness rather than to replace the ergometer with another focal dashboard (90, 94, 117).

This chapter describes the parts of the implementation that matter for interpreting the study in Chapters 4–6: the overall architecture, the sensor-fusion and stroke-arbitration design (which bears directly on data quality for RQ2), and the visualization mapping (which operationalizes the ambiguity-as-design-resource concept from Section 2.3 for RQ1). The underlying GPU fluid solver is a comparatively standard application of established real-time techniques; its full mathematical derivation, kernel inventory, and performance engineering are given in Appendix A for readers who want implementation-level depth, and are referenced only briefly here.

3.2 Design Rationale

Three high-level choices shape everything described in this chapter. Each follows from the ambient-feedback literature reviewed in Chapter 2 rather than being an incidental implementation detail, and each trades explicitness for openness in a way that the study (Chapter 4) is specifically designed to probe. They are also consistent with embodied-interaction perspectives in HCI, which argue that meaning in physically situated systems

emerges through bodily action and full-body engagement rather than through symbolic display alone (40, 46, 61, 66).

Floor projection rather than a screen or a worn display. Two alternatives were available and neither was chosen. The first is a screen. The closest comparable system, ARrow (62), is a tablet application: a coach points an iPad at the rower and stroke-phase and technique cues are composited onto the live video, with an option to stream that view to a second screen the rower can see. This is a foreground channel by construction — it is a display someone looks at, and its content is a prescription about where the body should be. RowSim instead projects onto the floor around the ergometer, which keeps the feedback in the visual periphery and is consistent with the tangible-media tradition of making architectural surfaces rather than screens the site of ambient information (64, 117, 130) — most directly with Dahley and colleagues' projection of digital state into architectural space as moving light (34) — and with the broader spatial augmented reality tradition of projecting computer graphics onto physical environments and surfaces rather than through a screen or a worn display (13, 95).

The second alternative is a head-mounted display, which is the obvious way to put graphics in a rower's own field of view and which rowing research has already taken up: Arndt et al. report better performance and experience in a head-mounted condition (6), and VR4VRT and the Brain-Row Challenge both build ergometer training on worn displays (74, 123). It was set aside here on practical rather than optical grounds. A head-mounted display can render peripherally, so the objection is not that it could not carry an ambient channel. It is that rowing at training intensity is a hot, sweating, high-amplitude whole-body activity: the torso swings through a large arc, the head accelerates with every drive and recovery, and a worn display adds mass to the head while sealing heat against the face. Sweat and condensation on the optics, slippage of the headstrap under repeated acceleration, and thermal discomfort inside an enclosure are all worse during vigorous exercise than during the seated tasks head-mounted displays are usually evaluated in, and each is a hygiene problem as well as a comfort one when the same device passes between participants. A floor projection has none of these properties: it touches nobody, adds no mass, and is unaffected by perspiration. Relatedly, much of the presence and immersion literature that motivates head-mounted designs (108, 109, 131) targets a construct RowSim does not primarily aim to elicit.

A continuous fluid simulation rather than explicit numeric or graph-based feedback. Rowing is a continuous, rhythmic activity; a real-time fluid field is continuously and smoothly responsive to forcing in the same way, making it a closer perceptual match to rowing rhythm than a discrete numeric read-out (split time, watts, stroke rate) refreshed

once per stroke. Numeric dashboards are excellent for tracking performance exactly, and that exactness is what commits them to focal inspection: a number has to be read to mean anything, and reading is not something done at the edge of attention (94). Section 2.2 gives the argument in full: symbolic displays resist peripheral uptake because reading is not something done inattentively, and in Buxton's terms a console is a high-bandwidth, bursty channel that must be sampled deliberately, whereas a continuous field is a low-bandwidth, persistent one that need not be (25). Fluid motion accordingly lets rhythm, continuity, and effort remain legible without requiring the participant to read and interpret a number. In the study the console was covered for the recorded segment (Section 4.3), so the persistent channel was tested on its own rather than alongside the bursty one.

Ambiguity as an intentional design choice, not a missing feature. Section 2.3 sets out ambiguity as a resource that invites interpretation rather than conveying a single correct reading, and identifies the specific tactic RowSim uses: imprecise representation, or what Gaver et al. call digital sfumato — blurring the usual precision of an electronic display so that the viewer must bring it into focus themselves (47).

The ergometer console is the precise display being blurred. RowSim's visualization mapping (Section 3.10) implements the blur directly: colour is composed from the state of the simulation rather than from any reported quantity, so no numeric value is legible on any visual channel and none is recoverable to the digit — a claim about legibility, not about the absence of coupling, and Section 3.10 sets out exactly which quantities do drive the display. This was a deliberate decision, not an omission of a numeric overlay — a fully legible readout was considered and rejected precisely because it would restore the definition the design is trying to reduce, and would collapse the interpretive openness that RQ1 and the interview protocol are designed to examine.

The same framing bears on what the simulation is for, and it is worth separating two things that could be confused. RowSim's *sensing* is not loose. Stroke events come from the ergometer's own instrumentation, arbitrated as described in Section 3.9, and the ingest path is measured in Section 3.11 at a median of 1.13 ms. What is deliberately not exact is the *fluid*: it is a controller, not a hydrodynamic model. It does not follow hydrodynamic truth, and it is not trying to. Its job is to behave plausibly under forcing — to shed, drift and decay in ways a viewer reads as flow — not to predict what water would do.

That is a design commitment rather than a concession, and it is the same commitment the colour channel makes. RowSim is not a replica of water and does not try to be. It is a departure from realism into an abstract register — a field that moves the way water moves, rendered in a magenta no river has ever been. What it serves is the abstraction, not the replica, and the two want different things from a simulation. A replica wants to survive

measurement. An abstraction wants to be read. Section 4.1 makes the same move in colour, and for the same reason: what is abandoned is resemblance, not rigour. Chapter 5 reports how participants read the result, including two elite rowers who identified specific departures from on-water behaviour and neither of whom described them as errors.

Table 3.1 summarizes these three choices alongside lower-level engineering decisions discussed later in this chapter, together with the alternative considered for each and why it was not adopted.

Table 3.1: Key design and engineering decisions, alternatives considered, and rationale.

Decision	**Alternative considered**	**Why rejected**
Floor projection	A screen the rower looks at (as in ARrow (62)), or a head-mounted display	Coaching overlays of the kind ARrow provides are read from a screen, in central vision; floor projection keeps feedback peripheral. A worn display could render peripherally but adds mass to a head that accelerates every stroke, and faces sweat, heat and slippage during vigorous rowing
GPU grid-based fluid simulation	CPU-side particle system	A continuous velocity field degrades gracefully into ambient motion rather than resolving into discrete, individually trackable objects, better matching the ambiguity goal (Section 2.3)
Ambiguous, multi-feature colour mapping	Explicit numeric overlay (split, watts, stroke rate)	Would collapse the interpretive openness that RQ1 and the interview protocol are designed to examine
Native Metal compute (4)	Cross-platform engine (e.g. Unity) or OpenGL/Vulkan	A general engine would have been viable; native Metal was chosen for direct control of command-buffer scheduling and kernel timing, which made the semaphore and ring-buffer reasoning in Section 3.11 easier to state. This is a project judgment rather than a demonstrated necessity
Centralized stroke ownership (Section 3.9)	Independent multi-sensor triggers	Prevents several sensing sources acting on the same physical stroke; note this governs interaction, not the derivation of the analytical stroke series (Section 4.6)

3.3 Physical Installation

RowSim is a spatial system, and the spatial arrangement is part of the design argument: the projection surrounds the ergometer on the floor rather than sitting in front of the rower, which is what places it in the visual periphery during the drive phase.

Two projectors cover the floor area, which is what makes the projected region wide enough to extend past the rower on both sides rather than forming a pool directly beneath the machine. A boat-shaped cutout is masked over the ergometer, so that projected light does not fall on the machine itself and the unlit region reads as a hull with the fluid field flowing around it. The silhouette is generated rather than traced: a per-fragment test decides whether a point falls inside the hull, narrowing the half-width from midships toward each end. The taper is deliberately asymmetric, finer at the bow than at the stern, and its proportions follow the plan view of a racing single scull, so that the length-to-beam ratio and the fore-aft asymmetry a rower would recognize survive into the projection (Figure 3.2). Appendix A gives the exponents. This is a deliberately low-technology solution to a problem that would otherwise require per-frame geometric compensation for the ergometer's silhouette, and both elite participants singled the resulting form out unprompted: P10 valued "the overlay of a simulated boat where the rowing machine is sitting," and P13 reported that "I like the boat a lot, because it does really give you a sense that you're in the boat."

Figure 3.3 shows the installation as deployed in the Graphics and Experiential Media (GEM) Lab, and Appendix A.8 collects further views of it. What the projection itself looks like is taken up in Section 3.10, where the palette variants are set out alongside the mapping that drives them (Figure 3.13).

Preparing the surface. Room lights were switched off for every session, so the projection had the room to itself. The floor it landed on is matte and near-black, built to absorb stray light from the overhead rigs and correspondingly reluctant to return any. Projected directly onto it, the fluid lost most of its tonal range before it reached the eye.

The fix is unglamorous and was arrived at by trial (Figure 3.1). Card panels went down first to test whether reflectance was in fact the problem, then progressively larger sheeting, and finally four white sheets taped together to cover the whole projection area. That is the surface every participant rowed over, and it is visible in Figure 3.3. What it recovers and what it cannot are taken up in Section 6.3, where they bear on the findings.

(a) Card panels **(b)** Sheet edge **(c)** Sheeted

Figure 3.1: Why the projection surface mattered, in a darkened lab. (a) Card panels laid on the bare floor, with the same projection falling on bare floor above them: legible on the card, a barely-there smudge on the floor. (b) The clearest comparison, one projector and one frame across two surfaces — saturated and detailed on the sheet at left, washed to a faint smear on the bare floor to its right. (c) Full sheeting, with the fine structure of the flow legible. The palettes here are development configurations and not the one participants saw; what the sequence documents is reflectance, not colour.

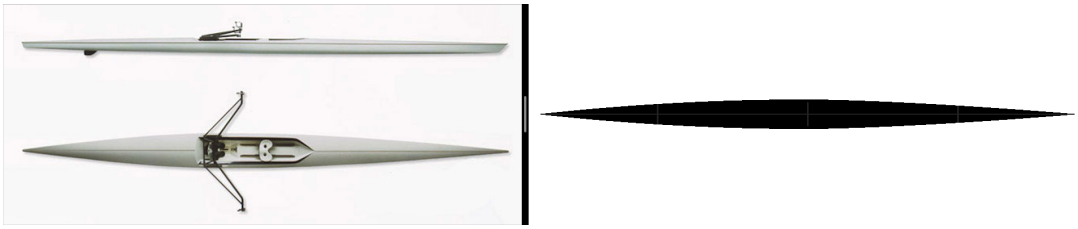


Figure 3.2: Derivation of the projection mask. Left, the reference used during development: profile and plan views of a racing single scull. Right, the resulting hull region, which the visualization pass uses as the unlit cutout and the solver uses as the partial-slip obstacle of Section 3.6. The shape is not a traced outline or a stored texture: it is an inside/outside test evaluated per fragment, so it costs no memory and scales with the configured boat size. **The edge shown here is the raw mask, before multisampling.** The inside/outside test is binary, so at simulation resolution the boundary is hard; what participants saw was resolved through the $4\times$ multisample anti-aliased (MSAA) colour target (Appendix A.2), which softens it. The visible stair-stepping is an artefact of this figure, not of the projection.

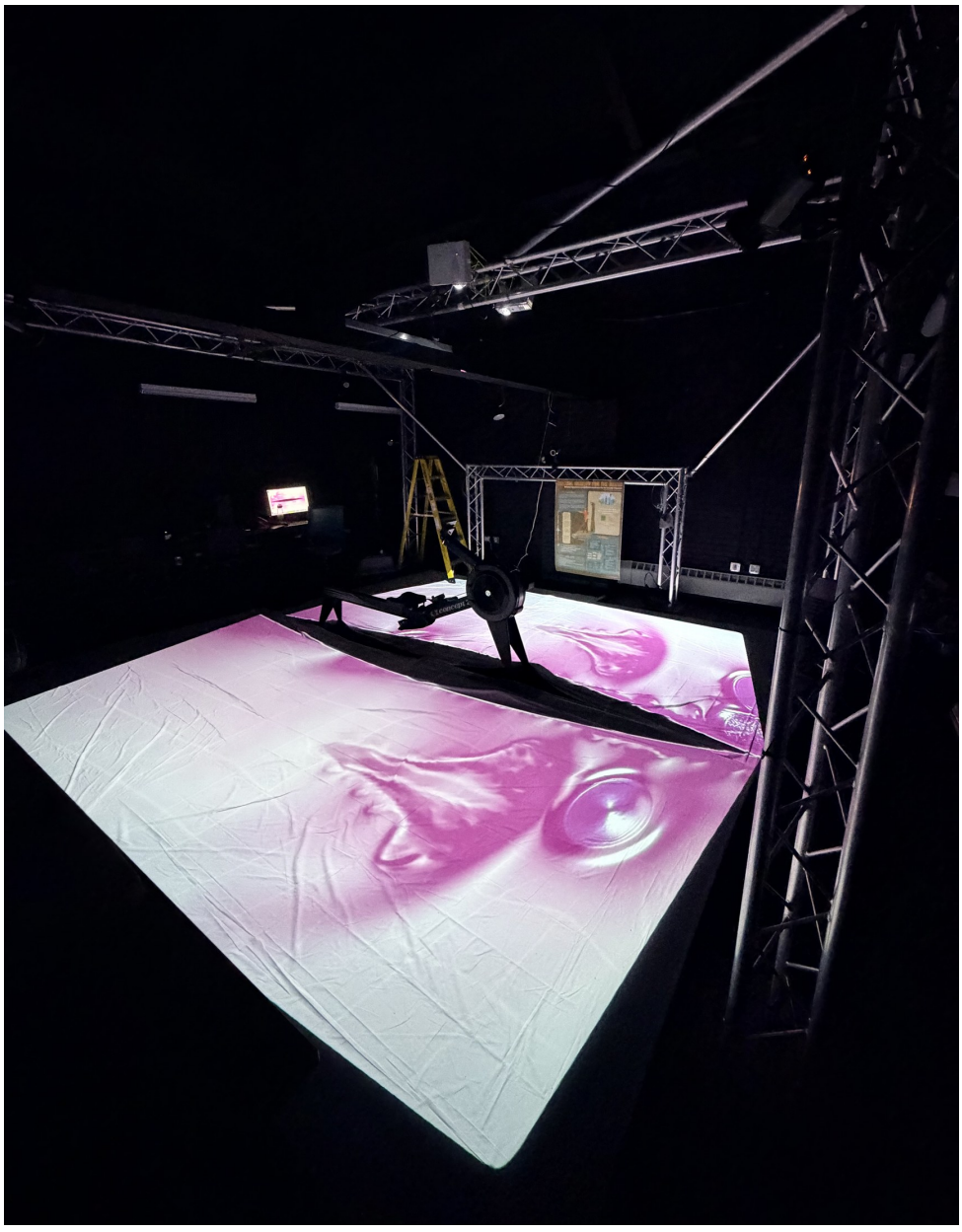


Figure 3.3: RowSim as deployed in the GEM Lab, photographed under session conditions: room lights off, with a white sheet laid over the floor as the projection surface. Overhead projectors place the visualization around the ergometer in the rower's lower periphery; a projection mask leaves a boat-shaped unlit silhouette around the machine. The sheet is load-bearing rather than incidental — Figure 3.1 shows why.

3.4 System Architecture

RowSim is written in Swift, with the simulation and visualization kernels in Metal Shading Language, and builds as a single Xcode project targeting macOS (the deployed platform) and iOS (a touch-only variant not used in this study). It depends on Apple's Metal, MetalKit and Core Bluetooth frameworks and on libmapper for the inertial transport; no game engine or third-party rendering library is used. Nothing in the design requires Apple's stack in particular — the solver is standard — but the choice is stated here because it constrains reproduction, and Table 3.1 records the alternatives that were considered.

The system separates orchestration from rendering. The macOS application host owns rowing state, sensor ingestion, calibration, PM5 coupling, and event arbitration. The shared renderer owns Metal resources, simulation textures, GPU pipeline states, and presentation. This dependency direction is deliberate: input providers feed an input router; the router forwards normalized events to the host controller; the host updates renderer state; and the renderer advances and presents the simulation. The renderer never reaches back into the application or into device-specific protocols (Figure 3.4, Table 3.2).

Table 3.2: Subsystem responsibilities in RowSim.

Subsystem	Primary responsibility	Research rationale
Input providers	Parse PM5, libmapper, and UDP signals	Allows device replacement without changing the study-facing event model
Input router	Fan in providers, emit typed events	Makes heterogeneous sensing auditable as one ordered interaction stream
Host controller	Maintain rowing phase, calibration, stroke ownership, mappings	Keeps physical interpretation inspectable and loggable on the CPU
Renderer	Advance simulation fields, present frames	Concentrates dense numerical work on the GPU
Visualization shader	Convert fluid, wave, bloom, and mask state into colour	Links numerical state to ambient feedback rather than numerical coaching

The platform split follows this separation. The macOS target contains the full research stack: PM5 Bluetooth Low Energy (BLE), native libmapper, UDP input, rowing physics, logging, and the shared Metal renderer. An iOS target reuses the shared renderer but supports direct touch input only, and was not used in the study reported here (Figure 3.5).

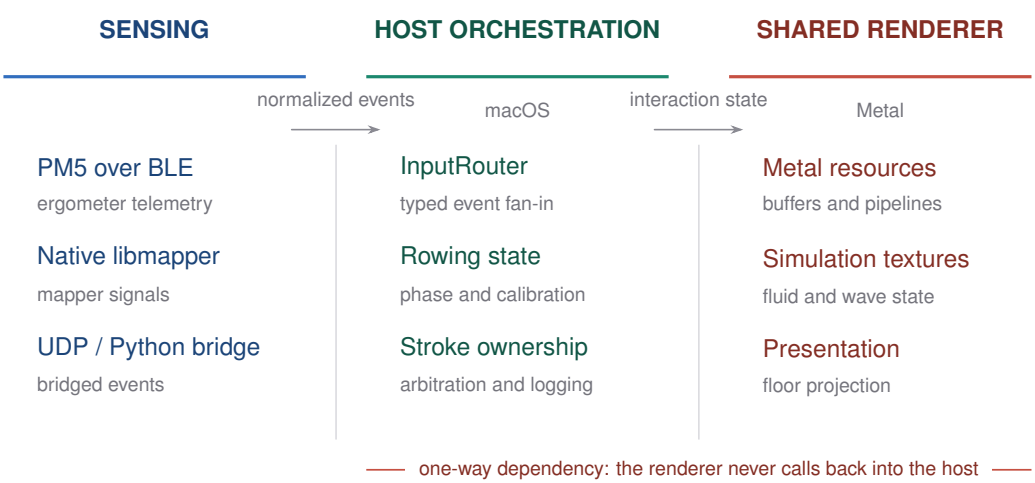


Figure 3.4: End-to-end RowSim architecture. Inputs are normalized before affecting rowing state, and the renderer consumes only host-level interaction state. The one-way boundary permits device substitution, and would permit replay from a logged event stream, without changes to rendering code.

The two sensing channels are complementary rather than redundant, and either will drive the system alone. The ergometer supplies stroke timing and the performance quantities; the handle-mounted inertial unit supplies what the ergometer cannot see — where the hand actually went, and how it was turned. Everything that responds to the shape of a stroke rather than to its magnitude comes from the inertial unit: the wake geometry traced from handle displacement, the ripples raised at the blade, and the caustic brightening that follows handle rotation. RowSim therefore runs in three configurations: ergometer alone, inertial unit alone, or both together. The study used both, which is the only configuration in which the display responds to timing, effort and gesture at once.

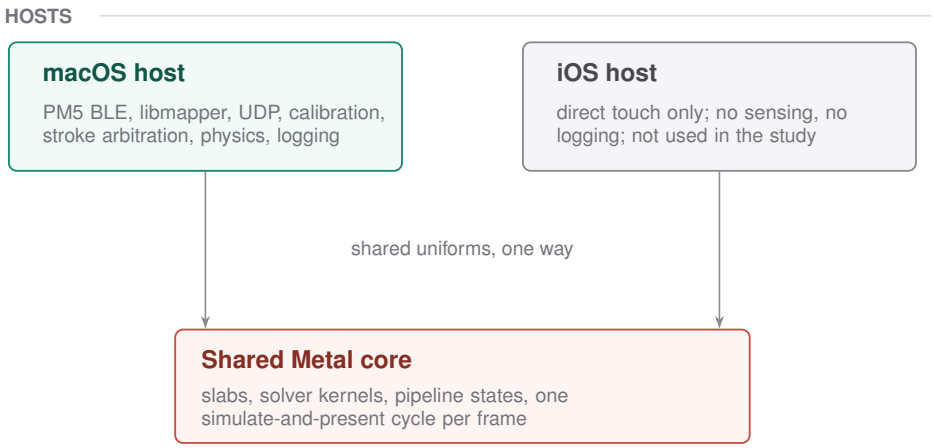


Figure 3.5: Two hosts over one renderer. The macOS target owns everything that makes a session interpretable — sensing, calibration, stroke arbitration, physics and logging — while the iOS target reuses the same Metal core with direct touch as its only input. The dependency runs one way in both cases: neither host is reachable from the renderer. Every session reported in this thesis ran on the macOS host.

3.5 Runtime Execution Model

Only the main thread may change simulation state. RowSim does not create a dedicated CPU render thread. Rendering is driven by the MetalKit view delegate: each display refresh invokes the renderer's draw cycle on the application main thread, where the CPU updates frame uniforms and encodes Metal commands; GPU execution then proceeds asynchronously after command-buffer submission. Two mechanisms bound CPU–GPU concurrency: a three-frame in-flight semaphore, which prevents the CPU from queuing unbounded GPU work, and a three-buffer uniform ring, so each frame writes a different uniform buffer than the one the GPU may still be reading. Sensor ingestion is asynchronous (native-mapper polling, PM5 BLE callbacks, UDP receipt each arrive on their own queue), but every event is normalized and re-dispatched onto the main thread before it touches rowing state. The central invariant — and the one that matters for data integrity — is that device callbacks may occur off the main thread, but rowing-state mutation, and therefore anything written to the session log, occurs on the main thread in a single, serial order. Figure 3.6 shows this arrangement: several concurrent producers on the left, one serialized point of mutation in the middle, and bounded CPU–GPU overlap on the right.

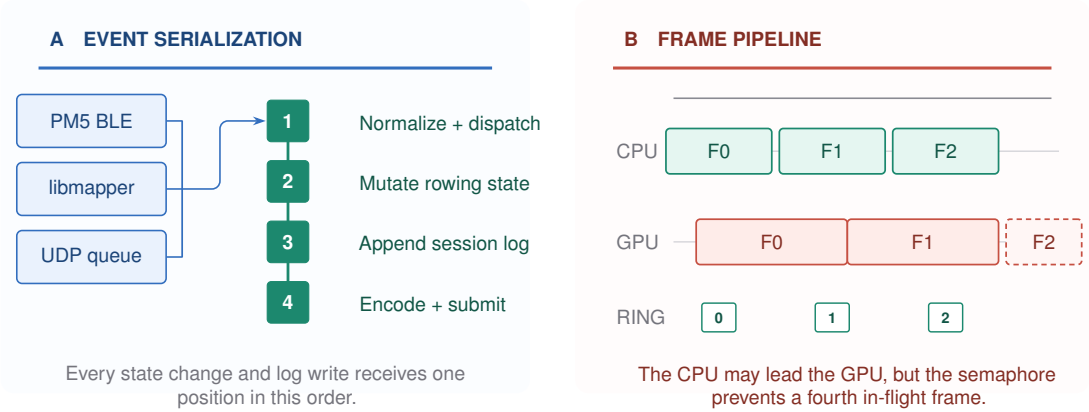


Figure 3.6: Runtime execution model shown as two related invariants. (A) Independent producers converge on one main-thread order for mutation, logging, and submission. (B) CPU encoding can run ahead of GPU execution, but three uniform slots and the semaphore bound the pipeline to three in-flight frames. Bar lengths are schematic, not measured durations.

3.6 Fluid Simulation Model (Summary)

It is more important to have beauty in one's equations than to have them fit experiment.

— Paul Dirac (37)

Dirac meant it as a claim about physics, where it is contested. It is quoted here because it names the trade this chapter makes without apology. RowSim's fluid answers to the eye rather than to the wind tunnel, and the space it works in is a genuinely interesting one: a two-dimensional surface embedded in a three-dimensional room, where the physical and the computational meet on the same plane. That surface admits things the room alone does not. A bag set down on the floor becomes an obstacle the water must part around. A body leaning across the projection displaces the flow. The rules are ours to write, and the interesting question is not how closely they track the Navier–Stokes equations but which departures a rower reads as water and which they read as a mistake.

RowSim's visual core solves a two-dimensional incompressible form of the Navier–Stokes equations in real time, using the operator-splitting decomposition — advection, forcing, projection, visualization — established for interactive graphics (22, 43, 113, 114). The equations are the real ones; what is approximate is the discretization and the iteration budget, which are chosen for a frame deadline rather than for convergence:

$$\overbrace{\underbrace{\frac{\partial \mathbf{u}}{\partial t}}_{\text{local}} + \underbrace{(\mathbf{u}\cdot\nabla)\mathbf{u}}_{\text{advective}}}^{\text{ACCELERATION}} = \overbrace{\underbrace{-\nabla p}_{\text{pressure}} + \underbrace{\nu\nabla^2\mathbf{u}}_{\text{viscosity}} + \underbrace{\mathbf{f}}_{\text{rowing}}}^{\text{FORCE PER UNIT MASS}}, \qquad \underbrace{\nabla\cdot\mathbf{u} = 0}_{\text{INCOMPRESSIBLE}} \tag{3.1}$$

where $\mathbf{u}$ is velocity, p is pressure, ν is viscosity, and $\mathbf{f}$ is the external forcing induced by rowing interactions. The rowing term $\mathbf{f}$ is the only place the rower enters: everything a participant did arrived in the simulation as a contribution to that one symbol, and everything they saw was the field's response to it. It is worth being exact about what Equation 3.1 does and does not license, because "fluid simulation" covers everything from a finite-volume solver to a shader that swirls. What runs here is a projection-method solve of the incompressible equations: each frame performs semi-Lagrangian advection, then force injection, then vorticity confinement, then a divergence computation, then a finite pressure Poisson solve — sixteen Red–Black Gauss–Seidel pairs on Apple Silicon, twenty Jacobi iterations on the fragment fallback — and then subtracts the pressure gradient to restore $\nabla \cdot \mathbf{u} = 0$. That is a real incompressible core, and the divergence-free constraint is enforced rather than approximated away.

Four things it is not. It is two-dimensional, not three. It is operator-split, so advection, forcing and projection run in sequence rather than as one coupled system. The pressure solve is truncated at a fixed iteration count rather than run to convergence, and semi-Lagrangian advection is numerically dissipative, so energy leaves the field faster than the equations say it should. That dissipation is the price rather than the point: what the scheme buys is unconditional stability, which is what allows one large fixed time step per frame instead of the small ones an explicit Eulerian advection would need to stay stable at this grid size (113). And the fields are half-precision `rg16Float`, sized for a 60 Hz floor projection rather than for accuracy.

Some of what a participant sees is therefore not hydrodynamics at all. Vorticity confinement (116) is a corrective force that puts back rotational detail the advection scheme removed; the ripple heightfield is a separate damped wave system; force caps, bloom and palette shaping are perceptual layers on top. The right position is to defend the incompressible core as what it is, and to claim nothing hydrodynamic for the layers above it. Pressure projection is solved with an in-place Red–Black Gauss–Seidel sweep on Apple Silicon and a ping-pong Jacobi sweep on the fragment fallback path, both implemented as GPU passes in the manner established by real-time GPU fluid work (31, 52); both branches apply the same stability clamps and iteration budgets, though they are not bit-identical and their convergence was not compared empirically (Appendix A, Section A.4). A multigrid pressure solver would converge in fewer iterations on large grids (79), but the fixed-iteration relaxation used here is sufficient at RowSim's grid resolution and keeps per-frame cost predictable, which matters more for a display whose latency budget is fixed by the refresh rate. Ripples are modelled as a separate damped wave field coupled back into velocity, and bloom/particle fields provide perceptual persistence rather than additional physical state.

The full kernel-by-kernel derivation, iteration counts, and numeric stability constants are given in Appendix A, Section A.3.

The hull is not a wall. One boundary condition is worth stating in the body of the chapter rather than leaving to the appendix, because it is a design decision expressed as physics. The solver carries an obstacle field, and obstacles are ordinarily no-slip: velocity inside them is forced to zero, so flow stops at the boundary and parts around it. The boat is deliberately exempted. Texels inside the hull are assigned a velocity proportional to the carrier flow rather than zero, tapering with distance from the edge and increased toward the bow and stern, and the tangential slip that results at the hull boundary is fed back as additional rotation so that the ends visibly shed.

The effect is that the hull does not read as an obstruction in a moving river. It reads as a body moving through still water, which is what a rower in a boat actually is. The simulation is computed in the boat's frame — the hull is fixed on screen and the water is given the relative motion — and the slip band is how a body that is nominally stationary in that frame is made to behave like one that is moving in the world's. A no-slip hull would have been more physically conventional and would have looked like a rock in a stream.

This matters for two reasons beyond appearance. It is a second instance of the chapter's general position, that the fluid is a controller rather than a model: the departure from the physically standard boundary condition is what makes the intended reading available, and the appendix records it as a departure. And it bears on a limitation identified in Section 6.1.6. Every session began from rest: the field was reset and carrier velocity was zero until the first stroke, so the motion a participant saw was motion they had produced. What the boundary condition changes is not where the flow starts but how it ends. Because the hull is given whatever carrier velocity the controller holds, the rower can add to that velocity and cannot subtract from it. Left alone it decays — the flow settles because nothing is driving it, not because the rower has stopped it. That is a defensible choice for a boat that is meant to be under way, and it is also why bringing the display to rest is something the rower waits out rather than something the rower does.

3.7 Per-Frame Render Graph

Figure 3.7 summarizes the fixed sequence of simulation and presentation stages executed each frame. The same conceptual graph runs on both supported architectures, but Apple Silicon executes the primary simulation with compute kernels, while the fallback path implements the core stages with fragment passes and omits the bloom-blur and particle stages (Appendix A).

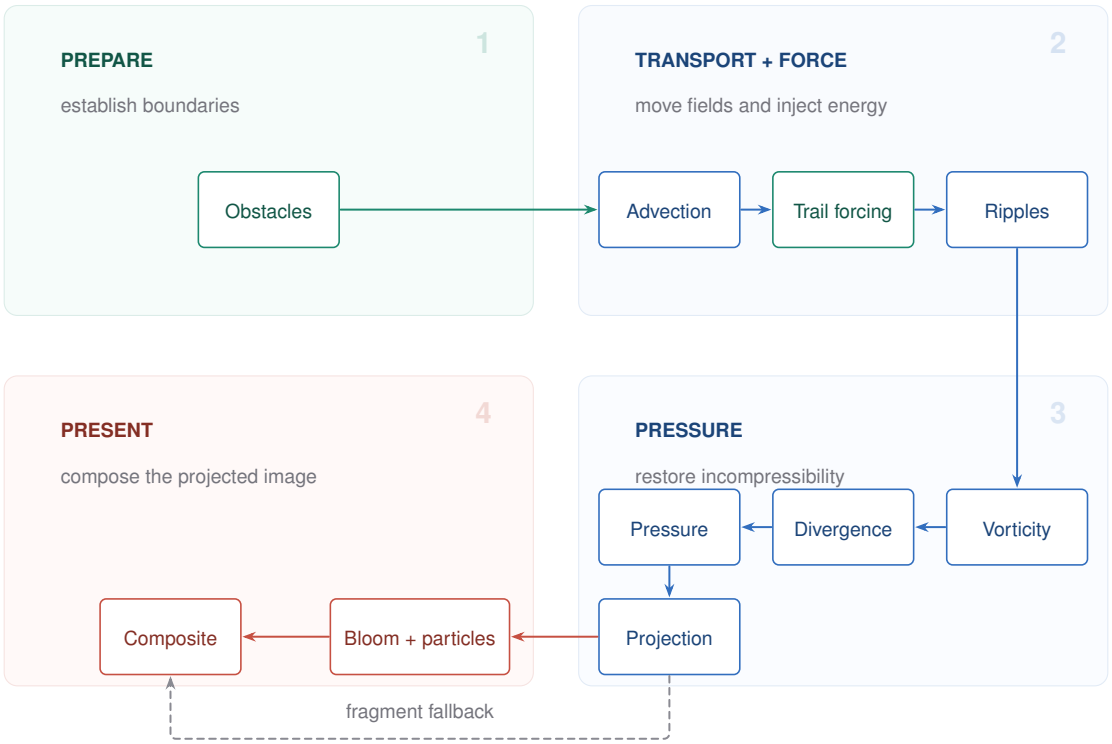


Figure 3.7: Per-frame render graph organized as four visual phases. Solid arrows trace the Apple Silicon pass order clockwise from preparation to final composition. The dashed fragment-fallback edge bypasses bloom and particles.

3.8 Sensor Architecture and Event Normalization

The input stack is designed around a transport-independent event model (Figure 3.8). PM5 BLE, native libmapper, and UDP/Python-bridge messages are all converted into typed events before reaching the host controller. The event vocabulary includes coordinates, gestures, acceleration, world acceleration, velocity, displacement, orientation, roll, azimuth, tilt, PM5 snapshots, and PM5 drive events. This fan-in architecture matters for reproducibility: the same rowing-state logic can be exercised by live PM5 telemetry, bridge telemetry, or synthetic events, and every session records which input path was active.

Transport is not meaning. The PM5 BLE client connects to the Concept2 PM5 service and subscribes to five characteristics: general status, additional status 1 and 2, stroke data, and additional stroke data (Appendix A, Table A.3). Drive events are deduplicated using a signature derived from stroke state, drive time, recovery time, drive length, and peak force, with an additional cooldown gate in the event sink to reduce duplicate drive-edge artefacts caused by asynchronous BLE characteristic arrival.

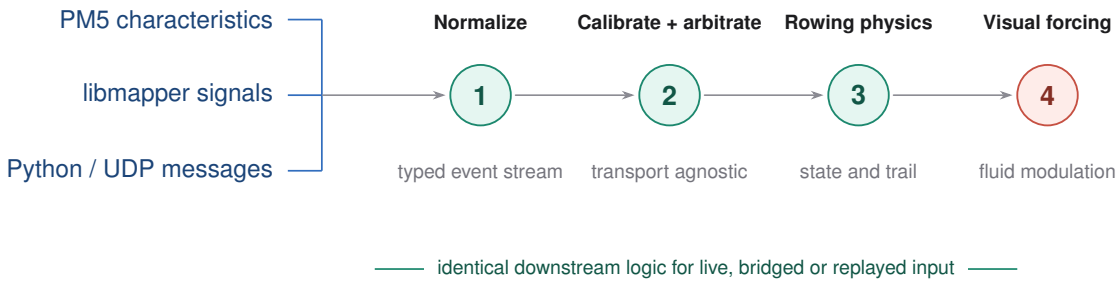


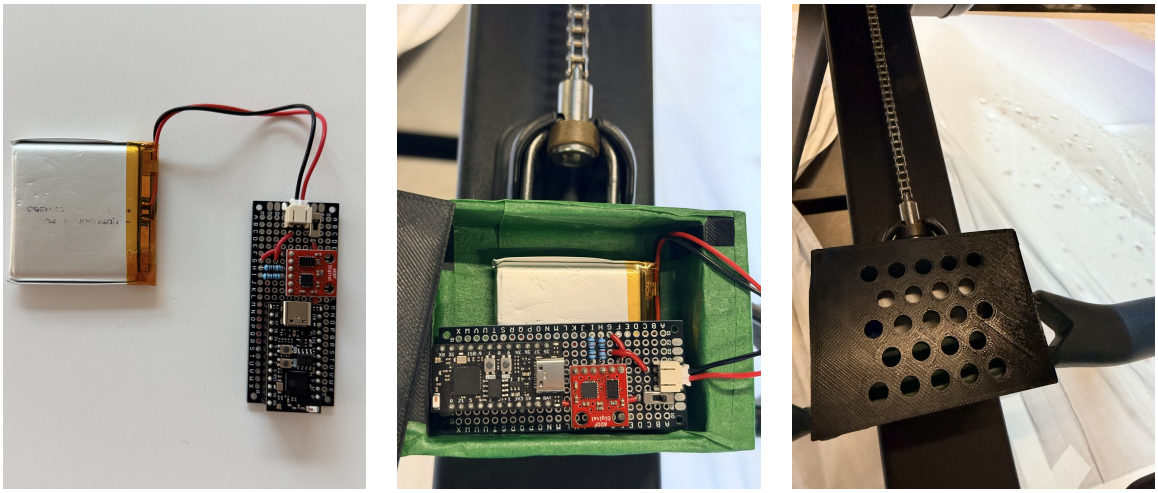
Figure 3.8: Sensor-fusion and interaction pipeline. Three transports become one typed event stream before calibration, stroke ownership, physics, and visual forcing. Downstream logic is therefore identical for live and bridged input, and the same boundary is what would allow a session to be replayed from its own event log without touching anything downstream — an affordance of the design rather than a mode that was built or used here.

3.9 Calibration, Stroke Ownership, and Rowing Physics

This section is deliberately more detailed than the preceding ones, because these design decisions determine what the logged performance data actually represent.

Calibration. Velocity and displacement handlers first collect a baseline (40 samples each) and subtract it, so a stationary sensor reads approximately zero. The inertial measurement unit (IMU) path uses hysteresis: it adapts its baseline slowly under low motion, engages rowing only after three consecutive samples exceed a motion or jerk threshold, and disengages only after five consecutive sub-threshold samples. This asymmetric engage/disengage design prevents visual and logging chatter when the rower is momentarily still or when sensor noise crosses a threshold once.

Stroke ownership. Stroke events can originate from PM5 drive detection or from the handle-mounted inertial unit (Figure 3.9), or, outside study sessions, from an automatic demonstration timer. The inertial unit contributes through two distinct paths rather than one: a motion-detection path that reads acceleration directly, and a velocity path that reads its integrated `worldFrame/velocity` stream. Both are the same physical sensor. The study ran PM5 alone or PM5 with the inertial unit; no third device was involved. Rather than allowing all sources to fire independently, RowSim centralizes ownership in two stages. First, arbitration: whenever PM5 is connected, PM5 drive events own stroke triggering and both inertial paths are suppressed. Second, a refractory guard, which is per-path rather than global — 0.4 s between admitted triggers on the inertial path, a 0.25 s cooldown on PM5 drive edges, and a 0.3 s/0.25 s drive–recovery debounce on the inertial velocity path. Because PM5 was connected in all seventeen sessions, the 0.25 s cooldown was the operative guard throughout this study and the other two never fired. This is a conservative design:



(a) Sensor and cell **(b)** Housed **(c)** Mounted at the handle

Figure 3.9: Handle-mounted IMU: (a) sensor, microcontroller, and cell; (b) assembled 3D-printed enclosure; and (c) mounting position at the chain end, clear of the participant's grip. The IMU is optional only as a stroke *trigger* — the ergometer can fire strokes on its own. It is not optional for what a stroke looks like: the wake geometry is traced from its displacement stream, and the caustic layer runs on its azimuth signal alone.

the ergometer's own stroke semantics take precedence whenever available. Because rowing performance and feedback systems depend on accurate stroke timing and force interpretation, explicit ownership is preferable to allowing multiple sensors to independently trigger the same interaction from one physical stroke (15, 49, 60).

Where the blade path comes from. Arbitration decides *when* a stroke fires; a separate mechanism decides what shape it draws, and it is the most direct action-to-picture coupling in the system. Once the handle-mounted inertial unit has collected a forty-sample baseline, the oar trails are no longer a procedural arc timed to the stroke. They are generated from the sensor's own `worldFrame/displacement` stream — the second integral of measured handle acceleration, baseline-corrected — mapped into screen space and injected as the trail geometry. This holds for every live stroke source, PM5 included: the ergometer says when the drive began and how long it was, and the inertial unit says what the hand actually did in between. The dependency runs the other way too — until the forty-sample baseline is collected the displacement path declines to draw at all, so the opening of a session has no trail geometry rather than a synthetic substitute. At the transport's sample rate that window is well under a second, so in practice it covers the first drive rather than several strokes. The wizard-of-oz demonstration timer is the only path that still draws a synthetic arc, and it was disabled in every session. What participants watched their wake trace, in other words, was a

measurement of their own handle path rather than an animation triggered by it. Figure 3.10 separates the two jobs a single stroke does.

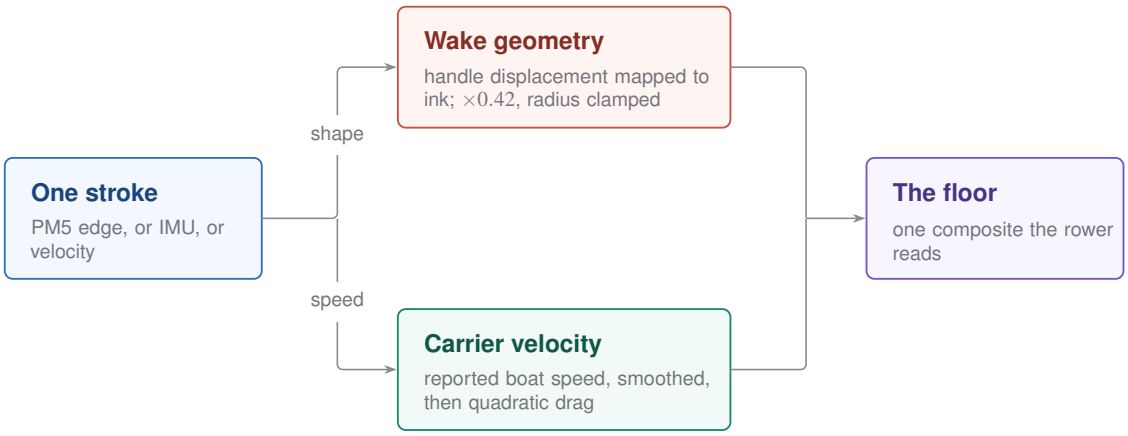


Figure 3.10: A single stroke does two separate jobs. Its shape becomes wake geometry, traced from the handle's measured displacement rather than drawn as a procedural arc; its magnitude becomes carrier velocity, smoothed from the ergometer's reported boat speed and thereafter decaying under drag alone. The two are computed independently and meet only on the floor, which is why the wake can answer a stroke that the carrier flow does not (Section 6.1.6).

What arbitration does and does not do. Precision matters here: the scope of this mechanism, because three different notions of "a stroke" are in play and conflating them would misdescribe both the system and the data.

1. *Raw telemetry.* The PM5 re-notifies its stroke-data characteristics several times per physical stroke. These notifications are written to the append-only event log exactly as received, deliberately: the raw stream is an audit record of what arrived, not a cleaned series.

2. *Arbitrated interaction triggers.* Ownership and the refractory guard govern which source is permitted to drive RowSim's visual response, so that a single physical stroke does not produce competing forcing events from the ergometer and the two inertial paths at once. This is what the arbitration guarantees, and it concerns the running system rather than the analysis.

3. *The analytical stroke series.* The stroke series used in Chapter 5 is derived from increments of the ergometer's own stroke counter, under the rule set out in Section 4.6 and validated there against the ergometer's independently reported stroke rate.

Arbitration therefore prevents *multi-sensor* double-counting at the point of interaction; it does not, and is not intended to, deduplicate the ergometer's repeated notifications inside

the raw log. The stroke counts reported later are the product of an explicit, documented derivation applied to that log — which is the appropriate way to obtain them, but should not be mistaken for a series that required no processing.

Figure 3.11 shows the resulting decision path. The design is deliberately conservative at two separate points: a single source is chosen before any event is admitted, and a refractory guard then applies to the admitted stream regardless of which source produced it.

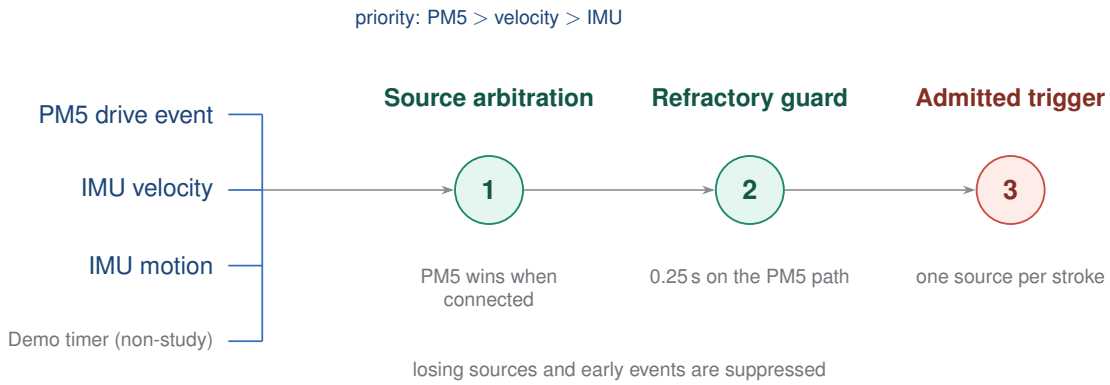


Figure 3.11: Stroke ownership and arbitration. One source owns triggering at a time, after which a refractory guard on the owning path admits at most one interaction trigger per physical stroke. The guard interval is per-path; with PM5 connected, as in every study session, it is 0.25 s. Raw PM5 notifications and the analytical stroke series remain separate concerns (Section 4.6).

Rowing physics. A simplified boat-velocity controller converts the active stroke source into continuous visual forcing. Two integrators are combined. The first is a threshold-gated propulsion and drag model driven by the inertial unit's integrated velocity stream. Carrier velocity is signed so that motion is negative and rest is zero, clamped to $[-v_{\max}, 0]$, so propulsion *subtracts* while drag adds back toward zero:

$$v_{t+\Delta t} = v_t + \Delta t\left(-F_{\text{drive}}(t) - k_d\, v_t |v_t|\right), \qquad F_{\text{drive}}(t) = \begin{cases} |\tilde{v}_x| \cdot \text{forceScale}, & \tilde{v}_x < -\theta_{\text{drive}} \\ 0, & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\tilde{v}_x$ is the baseline-corrected handle velocity and $F_{\text{drive}} \geq 0$. For $v_t < 0$ the drag term $-k_d v_t |v_t|$ is positive, so it opposes the motion, which is what Section 6.1.6 means by the flow decaying when nothing drives it. and a target-tracking controller that instead pulls boat velocity toward a PM5- or IMU-derived target velocity whenever one of those sources is active and recent, using a proportional response term. Equation 3.2 is that first path in isolation; when a recent PM5 or inertial target is available the target-tracking controller takes priority, per the stroke-ownership rule above. Both integrators are clamped to $[-v_{\max}, 0]$ and use a bounded per-frame time step to avoid integration spikes after a frame stall. This is not

a full rigid-body boat simulation — it is a controller that turns rowing effort into temporally coherent flow, trail geometry, and distance-related modulation, chosen so that the resulting visual response tracks stroke rhythm closely enough to support RQ1 without requiring a physically validated hydrodynamic model.

Table 3.3 lists the key numeric constants governing this behaviour, since they materially affect how sensor input is translated into both the visual response participants see and the interaction log used for RQ2.

Table 3.3: Key sensor and physics constants relevant to data interpretation. The physics values — drag, force scale, drive threshold and speed clamp — are recorded in all seventeen session manifests and were identical across them; they are the constants of the boat controller in Equation 3.2. The refractory intervals, calibration window, iteration counts and frame rates are source constants rather than snapshot fields, the last two selected at compile time by architecture.

Constant	Value	Relevance
PM5 drive-edge cooldown	0.25 s	Minimum spacing between admitted PM5 stroke triggers; the operative guard in every study session
Inertial source refractory	0.6 s	Minimum interval between inertial strokes; recorded in every session manifest
Unified refractory, inertial path	0.4 s	Cross-source guard applied on top of the above, so 0.6 s binds; suppressed while PM5 owns triggering
Quadratic drag coefficient C_d	0.04	Sole decay term on the *carrier* velocity between strokes; the velocity field has its own per-frame decay (Section 6.1.6)
Handle-speed to force scale	2.5	Base propulsive gain, blended with PM5 watts at runtime
Drive-phase threshold	0.15	Minimum calibrated $\lvert v_x \rvert$ counted as a drive
Carrier speed clamp	15.0	Hard ceiling on carrier flow magnitude
Velocity / displacement calibration window	40 samples	Baseline collected before a session's velocity/displacement readings are trusted
IMU consecutive samples to engage / disengage	3 / 5	Hysteresis band preventing spurious engage/disengage transitions
Pressure-solve iterations (Apple Silicon / fallback)	16 / 20	Affects visual smoothness only, not logged rowing data
Target frame rate (Apple Silicon / fallback)	60 / 30 fps	Affects visualization responsiveness; selected at compile time by architecture

3.10 Visualization and Ambient Feedback Mapping

The final visualization pass renders a full-screen quad that samples velocity, ripple height, bloom, particles, background, and obstacle textures in parallel (Figure 3.12). Rather than presenting split time, watts, or stroke rate as text, RowSim maps rowing rhythm and force into fluid motion, wake persistence, glow, ripple propagation, and colour. Colour in particular is deliberately not a read-out of any reported quantity. RowSim carries eight selectable palette variants — prismatic, watercolour, aurora, oil slick, thermal, fractal, mandala paisley and warped melting — which derive colour from five local flow features computed per fragment: flow speed, flow direction, curl magnitude, divergence, and a combined turbulence term. A ninth entry in the same menu is not a palette at all: it swaps the entire renderer for a separate implementation modelled on Pavel Dobryakov's widely copied WebGL fluid simulation (39), cited here because that is where its look comes from. Under those variants a single wake carries structured hue variation rather than a flat tint. A separate mapping from flow speed to colour is available independently of them. **None of that was used in the study.** Both the palette path and the speed-to-colour mapping sit behind runtime flags that were false in all seventeen sessions (Section 4.1), so what participants saw was one fixed hue whose opacity against the background follows local flow speed, with a depth-shading pass — refraction against the ripple height field, Fresnel and specular response, and slight chromatic aberration — supplying the remaining structure. Bloom and particles are additional compute-path overlays, while the background and obstacle textures establish the ground and unlit cutout. This design draws on visualization research showing that mapping continuous data to perceptually effective visual channels is itself a substantive design problem (29, 85), applied here toward ambiguity and legibility-at-a-glance rather than toward precise value recovery.

What is mapped, and what is not. Physics provides continuity; craft provides legibility. "Ambient" is used loosely enough that this deserves an exact statement, because the honest answer is not that nothing is mapped. Three ergometer quantities drove the display in every session. Watts scaled the brightness of the injected ink. Peak force scaled the impulse that splat delivered into the velocity field. The ergometer's reported boat speed set the carrier velocity of the whole field, and watts additionally blended into the propulsive force scale of the boat controller.

Two qualifications keep that list honest. The first is that the wake geometry in a study session came from the displacement path (Section 3.9), which attenuates what it inherits: ink intensity is scaled by 0.42 and the force multiplier by 0.55. The splat radius is a special case worth stating because it looks like a fourth coupling and is not. Two pieces

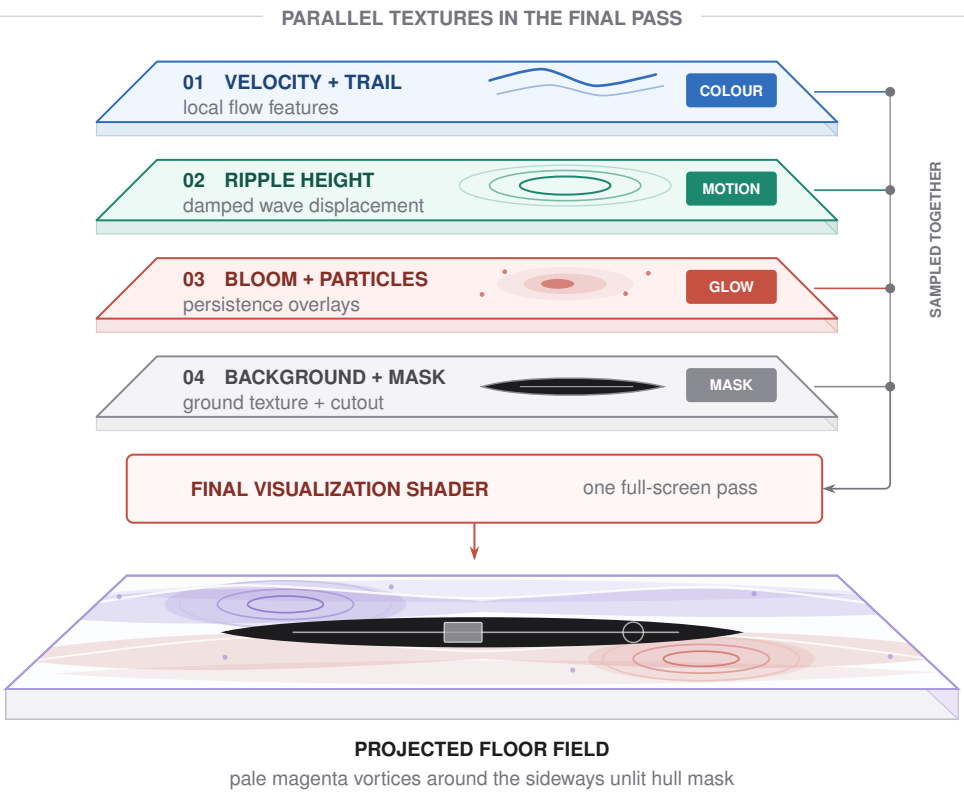
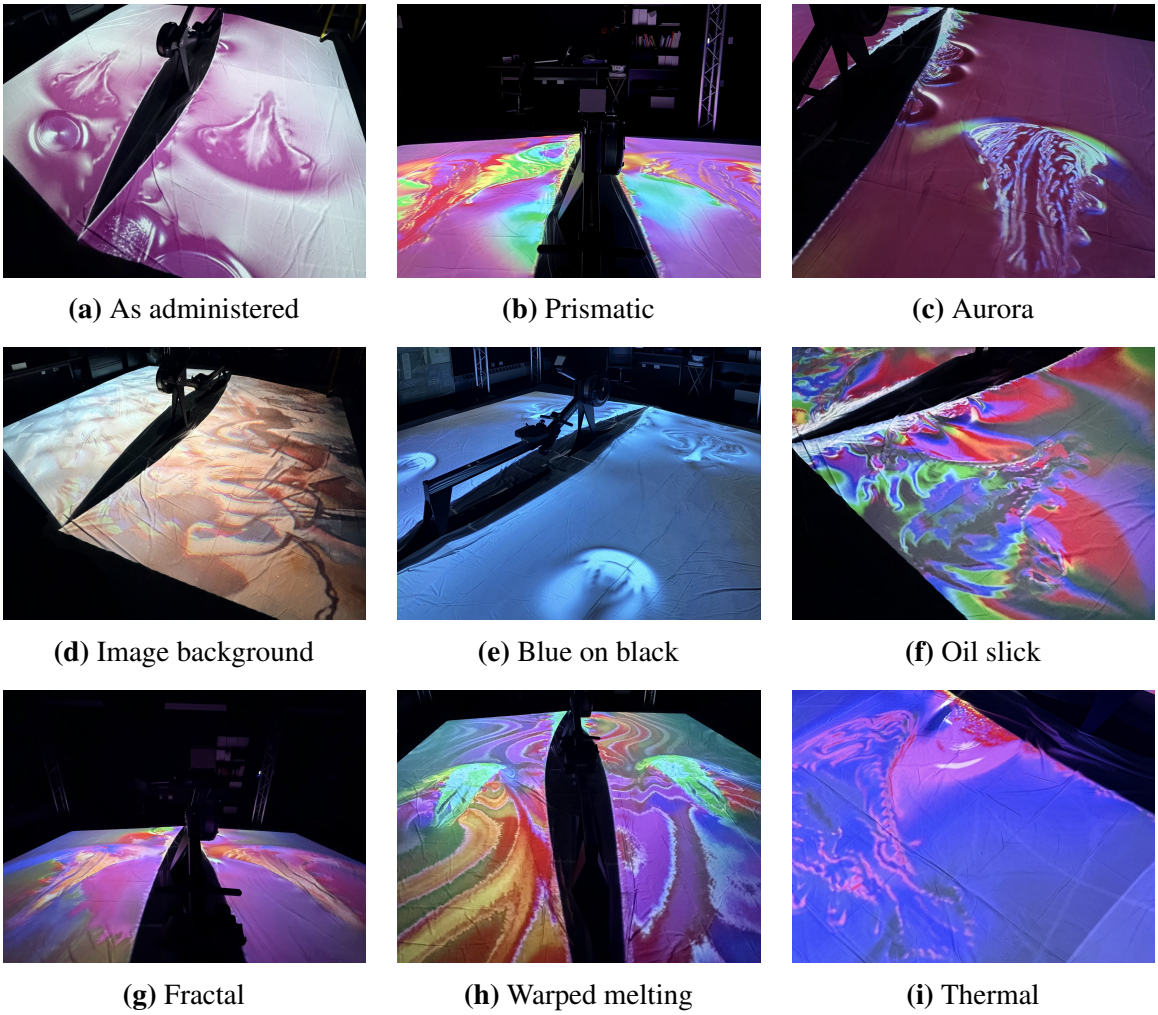


Figure 3.12: Visualization composition pipeline. Four registered state textures feed one full-screen shader pass: flow drives colour, ripple height supplies motion, bloom and particles add persistence, and background and obstacle textures establish geometry. The exploded stack denotes parallel inputs, not render-pass order. Two of the four ran in a reduced form during the study, and the figure shows the system's full capability rather than that configuration: the palette variants that would have derived hue from five local flow features were disabled, so layer 01 contributed one fixed magenta whose opacity follows flow speed, and the particle overlay in layer 03 was off in all seventeen sessions, leaving bloom alone to supply persistence. What reaches the floor is a single composite in which the contributing channels are no longer separable — no channel decodes to a reported number, which is the property this section argues for.

of code write that radius in sequence. The ergometer path computes it from power as $\max(80, \min(200, 120 + 0.1\,W))$, which for any positive wattage is at least 120 and rises with effort. The displacement path runs afterwards and overwrites it with $\min(r, 95)$, capping whatever it inherits at 95. Because the first value is always above the cap, the second always returns 95 — so the number the shader received was 95 on every stroke of every session, and splat radius did not vary with effort at any point. The two couplings that did survive are real but weaker than the raw mappings suggest. The second is that the enable flags in the manifest record which mappings were *permitted*, not which code paths ran. Two further mappings sit behind those same flags — stroke rate to the density of points along the trail, and peak force to the amplitude of a sinusoidal oar sweep — but both live in the procedural



(a) As administered **(b)** Prismatic **(c)** Aurora

(d) Image background **(e)** Blue on black **(f)** Oil slick

(g) Fractal **(h)** Warped melting **(i)** Thermal

Figure 3.13: Nine RowSim visual configurations under comparable rowing input. Panel (a) is the configuration used in every study session; panels (b)–(i) demonstrate the configurable palette and background mappings described in Section 3.10. Colour was held fixed during the study so it could not become an experimental variable.

stroke generator, which is reached only in wizard-of-oz mode. That mode was off in all seventeen sessions, so neither mapping was active despite its flag reading true.

The wider design space, and what was left in it. The mappings above are the ones that reached the study. They are a selection from a larger set built during development, and the rest is worth recording, partly because a reader of the source will meet it and partly because what was rejected states the design position more sharply than what was kept.

Three couplings were explored to the point of running continuously, with no flag to switch them off, and were kept because they act on how the display *feels* rather than on what it reports. Sustained rhythm is the input to two of them, and they read it differently. A sliding window over the last eight strokes yields a momentum term from the window's length and a

consistency term from the variability of the intervals within it. Trail ink brightness responds to momentum alone, rising about 39% over a full window: a rower who simply keeps going, however raggedly, gets a brighter wake. The bloom overlay responds to the *product* of the two, so it additionally requires the strokes to be evenly spaced; at its ceiling it is 63% brighter and warmed roughly halfway from the configured magenta toward gold. Both reset after four seconds without a stroke, and the boosted ink is then scaled by 0.42 on the displacement path exactly as watts-driven ink is. The third is geometric rather than luminous: each trail splat is stretched along its own direction of travel relative to the carrier flow, so dabs read rounder when boat and stream match and more streaky when the rower is coasting, slowing or pulling against the current. None of the three is recorded in the session log, and their status in this thesis is accordingly that of design description rather than measurement.

A second group was built and deliberately switched off. RowSim can render distance as a numeral drifting in the flow itself rather than in a corner overlay, draw a finish line across the projection at a target distance, and fire confetti across the floor when the target is reached; all three ship disabled and were disabled throughout. They are the explicit, goal-directed feedback the study exists to do without, and having built them makes their absence a choice rather than an omission. Section 4.1 explains why the task had no target at all, which would have made a finish line meaningless in any case.

The through-line is a distinction the rest of this chapter depends on. A display can be tightly coupled to performance and still refuse to report it. Effort brightens the wake, rhythm warms the room, and coasting draws the dabs out — but no quantity is named, no threshold is crossed, and nothing arrives to say well done.

What was excluded belongs to a different category. No quantity was rendered as a symbol, none was mapped onto the fluid's hue, and none is recoverable as a value — a viewer cannot read watts off the brightness of a wake, because that brightness is simultaneously a function of how recently the stroke landed, how the field happens to be advecting beneath it, and where the bloom sits in its own decay. The distinction the design actually rests on is therefore not between coupled and uncoupled, but between magnitude legible *as intensity* and magnitude legible *as quantity*. Section 2.3's formulation is the precise one and survives unchanged: no single quantity owns a single visual channel. Section 4.1 records the flag states, and Section 6.1.6 returns to the one coupling that participants noticed was missing.

This design operationalizes the ambiguity-as-a-resource position from Section 2.3: the palette functions deliberately avoid a one-to-one mapping between a single numeric quantity (e.g. instantaneous power) and a single visual channel (e.g. brightness), instead composing colour from the state of the simulation as a whole, so that no single "correct" reading of the display exists. The intended trade-off is exactly the one identified in that literature

— openness to interpretation, at some cost to precision for viewers who want an explicit numeric readout — and the interview protocol (Chapter 4) is designed to surface where individual participants land on that trade-off.

3.11 Performance Engineering and Empirical Validation

RowSim's performance strategy keeps dense field operations on the GPU, avoids needless CPU–GPU synchronization, and bounds frame latency through the three-frame semaphore and uniform-buffer ring described above (Appendix A gives the full memory-bandwidth and branch-divergence analysis).

The end-to-end delay between a physical stroke and the corresponding change in the projected image decomposes into three intervals. The first is measured below from the session logs; the second and third are not, and the distinction matters for what this chapter can claim.

Ingest latency, measured. Every sensor record in `raw/events.jsonl` carries both the timestamp at which the sample originated and the timestamp at which the host ingested it, so this interval is recoverable from the sessions already collected without additional instrumentation. Across all seventeen sessions and 7,301,229 sensor events, the median ingest latency is 1.13 ms, with a 95th percentile of 2.61 ms and a 99th percentile of 3.64 ms. Per-session medians range from 1.08 ms to 1.17 ms, so the figure is stable across deployments rather than an average over heterogeneous conditions. Broken down by stream, the PM5 paths are comparable to the inertial ones: `pm5/general` has a median of 0.65 ms and `pm5/drive` 1.43 ms, against 0.07–1.79 ms across the inertial streams, with the ordering across those streams reflecting the sequential order in which a single sample's derived quantities are emitted rather than any difference in transport cost.

For context, RowSim renders at 60 Hz, so one frame is 16.7 ms. Ingest therefore consumes roughly 7% of a single frame interval at the median and under 22% at the 99th percentile (Figure 3.14): it is not where any perceptible delay in this system originates.

What remains unmeasured. Two intervals are not established by these logs. *GPU frame time* — the interval between scheduling and completion — is available from the existing command-buffer completion handler but was not logged during the reported sessions; it should be summarized as a distribution rather than a mean, since occasional long frames matter more perceptually than the average. *State-to-photons delay* — the residual between state mutation and the frame reaching the projector, including projector input lag — cannot be recovered from software timestamps at all and requires an external measurement, such as

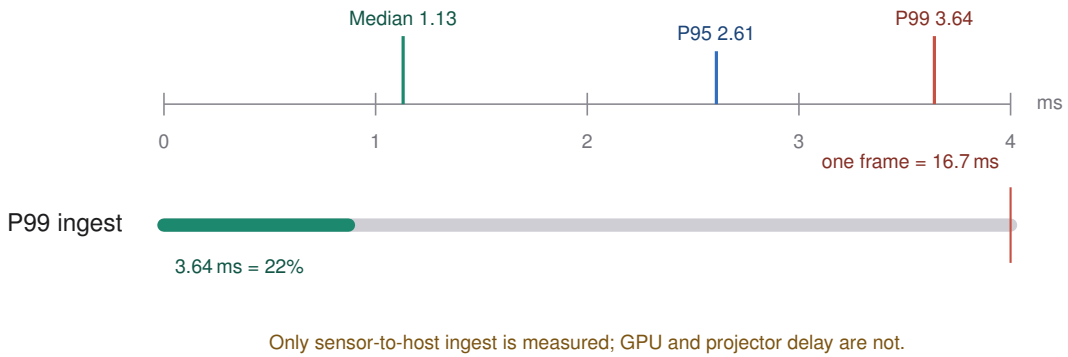


Figure 3.14: Measured ingest-latency percentiles relative to the 16.7 ms interval available at 60 Hz. Even the 99th percentile occupies under one quarter of a frame; this comparison does not include GPU execution or projector input lag.

a high-frame-rate recording capturing both the handle and the projected surface in one shot with the offset counted in frames.

This matters because feedback latency is not incidental to what this thesis studies. RQ1 concerns perceived coupling between rowing action and visual response, and perceived coupling is what end-to-end delay would degrade. The measurement above removes the sensing path from suspicion; the display path remains an argument from architecture, and is restated as such in Section 6.3.

3.12 Reproducibility, Assumptions, and Limitations

RowSim depends on a Metal-capable Apple platform, Bluetooth permission and PM5 availability for native ergometer telemetry, and a libmapper-compatible interface or UDP bridge for middleware-driven input. Runtime configuration affects behaviour — input mode, mapper interface selection, and bridge host/port settings all vary session to session — so each session's manifest records the input path, the host, the operating-system version and the application version. The architecture branch (Apple Silicon compute path vs. fragment fallback) is *not* a recorded field; it is determined at launch by device capability and is therefore recoverable from the recorded host rather than read directly. The two branches are numerically related but not identical (Appendix A).

The simulation itself is a perceptual, real-time approximation: semi-Lagrangian diffusion, finite pressure iterations, half-float precision, and branch-specific solver behaviour mean RowSim should not be interpreted as a physically validated water simulator. Its reproducibility claim is correspondingly narrow: the system records enough about each session — runtime configuration, a software snapshot giving platform, host, operating-system and

application version, and the full input stream — to make the path from rowing telemetry to projected feedback auditable after the fact. The architecture branch is not among those fields and is inferred from the recorded host, as stated above. Whether a re-run reproduces a session frame-for-frame was not tested, and GPU scheduling makes exact determinism an assumption rather than a verified property.

3.13 Summary

The RowSim implementation combines a modular sensing architecture, a main-thread state-mutation invariant, a bounded CPU–GPU execution model, a Metal-based fluid and wave simulation, and an intentionally abstract visualization pipeline. Its architecture supports the thesis's research questions by making rowing activity visible as continuous ambient motion rather than as additional numerical feedback, while keeping the sensor-to-log pipeline auditable enough to support the performance analysis in Chapter 5.

Chapter 4

Methodology

4.1 Research Design

This is a laboratory-based, mixed-methods study of RowSim with seventeen rowers. Each participant completed one open-ended rowing segment with the ambient visualization active, followed by two questionnaires and a semi-structured interview. Quantitative logging ran throughout.

The overall shape of the work — building a functioning artefact in order to ask a question that could not be asked without it, then studying how people respond to it — places it within the research-through-design tradition, where the artefact is a vehicle for inquiry rather than only an engineering deliverable (137, 138). RowSim's value is not that it is the optimal ambient rowing display, but that building and studying it makes a specific design position concrete enough to examine empirically.

Two features of the design are deliberate and shape everything that follows.

An open-ended rowing task

Participants were given no performance target, no duration to sustain, and no instruction to row continuously. The session script (Appendix B.4) reads: *"There is no right or wrong way to interact with the system. Please row naturally and focus on your experience,"* followed by *"You may start rowing whenever you are ready."* Participants were told they could pause, rest, or stop at any point. The researcher was instructed not to coach, not to give performance feedback, and not to ask questions during rowing, intervening only for safety or equipment issues. Before the questionnaires, participants were directed to answer with respect to the experience and the feedback rather than to their own fitness or athletic ability.

This is not an absence of experimental control but a match between the protocol and the artefact. RowSim's design position is that a display need not prescribe what the user should do (Section 2.3); a protocol that prescribed a target rate or a fixed piece would have supplied externally exactly the specification the display withholds, and would have measured

compliance with the instruction rather than response to the system. The task was left open so that what participants did with the display would be theirs.

The measurement consequence is substantial and favourable. Because nothing was mandated, everything recorded is voluntary behaviour: how hard a participant rowed, at what rate, for how long, and whether they stopped are all choices rather than compliance. Time spent rowing and the number of self-initiated pauses therefore function as behavioural indicators alongside the self-report instruments, and Section 5.3 treats them that way.

A single fixed visual configuration

Every participant saw the same visualization, in the same configuration, with no per-session adjustment. RowSim exposes a large runtime configuration surface (Appendix A), including several options that would have varied the display's appearance: automatic cycling of the fluid colour, automatic cycling of the background colour, a multi-colour "psychedelic" mode — a maximalist treatment of colour, with several variants, and a setting that maps flow speed onto hue. Every one of these was disabled for every session. The recorded colour configuration is identical across all seventeen manifests: fluid colour fixed at RGBA $(0.581, 0.088, 0.319, 1)$ — a deep magenta — against a white background at $(1.0, 1.0, 1.0, 1)$, with the colour-cycling, psychedelic-mode and speed-to-colour flags all set false.

The converse is worth stating with equal precision, because a reader who opens the repository will find it and because it bears directly on what the display was. Three settings were changed from their defaults and no others: the fluid colour, the background colour, and RowSim's own on-screen numeric overlay, which was switched off (Section 4.3). Everything else ran as the source ships it. That includes the four flags coupling ergometer telemetry to the wake and the two coupling it to the physics; all six read true in all seventeen manifests. Section 3.10 sets out which of them actually took effect on the code path the study ran, and which did not. What was held fixed was colour, which is why colour could not become an experimental variable; what was left at default was the effort-to-intensity coupling, which is the display's whole mechanism and was never a candidate for removal.

Two decisions are embedded there. The first is that colour is held constant so that it cannot become the study's operative variable. Colour is among the most immediately salient properties of a projected display and among the easiest to form a preference about; had it varied between participants, or cycled within a session, differences in reported experience could not have been separated from differences in colour preference, and the study would have become an experiment about palettes. Holding it fixed makes every participant's account an account of the same artefact.

The second is that the fixed colour is deliberately not the colour of water. RowSim does not attempt photorealistic water, and the choice of magenta rather than blue states that commitment in the most visible channel available. The system's fluid dynamics are legible as flow; its colour declines the invitation to be read as a literal river. This follows directly from the design position of Section 2.3: a representation that announced itself as realistic would invite evaluation against realism, whereas one that is plainly not a photograph of water asks to be read for what it does rather than what it depicts. Section 5.8 reports how participants responded, including one who articulated the rationale unprompted and better than the design documentation had.

One mode deserves separate mention. RowSim can drive hue from flow speed, which would have made colour a direct read-out of a single performance quantity. That mode was off in all seventeen sessions, as were the eight palette variants that would otherwise have derived hue from local flow features. The fluid's hue therefore carries no performance information and does not vary: what participants had to work with is the motion of the fluid itself (Section 3.10). The bloom overlay is the one exception and is not held fixed in the same way — steady rhythm warms it toward gold, as Section 3.10 sets out — so the claim here is that no reported quantity was mapped to hue, not that nothing on screen ever shifted colour. That shift is slow, applies to the glow rather than to the fluid, and was not treated as a colour condition. One further flag deserves the same treatment, because a reader looking at the source will find it and wonder. RowSim includes a *wizard-of-oz* mode that drives the display from a synthetic stroke timer rather than from the rower, so the system can be demonstrated without an ergometer. It was disabled in every session. Every stroke that moved the fluid during this study came from the person rowing.

The study follows principles consistent with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) (27), and was reviewed and approved by the Dalhousie University Social Sciences and Humanities Research Ethics Board under delegated review prior to recruitment (file 2025–8251; the letter of approval is reproduced in Appendix B.6).

4.2 Participants

Seventeen adults ($\geq$ 18 years) took part: twelve casual rowers and five elite rowers, characterized group by group in Table 4.1. The approved protocol allowed for up to twenty-four participants in two groups of twelve; recruitment concluded with seventeen completed sessions, and every analysis in this thesis uses that final sample. Group membership was

determined at recruitment screening and corroborated by the objective background items in the demographic questionnaire.

Why seventeen. Seventeen is where recruitment ended, not where it was aimed. The design targeted twenty-four participants in a balanced twelve-and-twelve split of casual and elite rowers, which would have given the group comparison equal arms. The casual arm filled; the elite arm did not, because the population is small, seasonal, and training on a schedule that leaves little room for an hour in a laboratory. What the study has instead is twelve casual and five elite rowers, and the consequences run through Chapter 5: the comparison is unbalanced, the elite arm is the one that limits it, and every group claim is reported with that asymmetry visible rather than averaged away.

That the achieved number is nonetheless defensible rests on information power rather than on saturation (75). Malterud and colleagues argue that the sample a qualitative study needs shrinks as the information each participant carries grows, and they make that a function of five things: the aim of the study, sample specificity, use of established theory, quality of dialogue, and analysis strategy.

Four of the five point toward a small sample here. The aim is narrow — one system, one session, one question about attention. The sample is specific: rowers rather than a general population, spanning a competence range chosen so the elite and casual accounts have cases on both sides. Established theory frames the reading in advance, from the ecological account of self-motion and from calm technology. And the dialogue was dense, because each interview followed a full instrumented session, so participants described something they had just done rather than recalling a category of experience.

The fifth dimension pulls the other way, and it is the one that governs what this study can claim. Malterud is explicit that an exploratory cross-case analysis needs more participants than an in-depth reading of a few narratives, and reflexive thematic analysis across seventeen accounts is cross-case work. Seventeen is therefore not a comfortable number on that dimension; it is the point where four favourable conditions meet one unfavourable one. The consequence is stated where it bites: the thematic claims in Chapter 5 are reported with their contributor counts so a reader can see how much of the sample each rests on, and the group comparison in Section 5.4 is the part of the analysis most exposed by it.

Saturation is not claimed, and Braun and Clarke argue it should not be: information redundancy is incoherent as a stopping rule for reflexive thematic analysis, because meaning is generated in the analysis rather than exhausted by the data (21). What can be said is narrower, and it is checkable against the corpus. Seventeen sessions produced the six themes reported in Section 5.6, and every one of them is carried by participants from the first

fourteen sessions; the last three contributed to existing themes rather than opening new ones. That is a property of this dataset, not a guarantee about rowers in general.

What the two labels mean here. Both are operational categories defined by the recruitment criteria below, not judgements of skill. *Elite* means meeting the training and recent-competition criteria stated for that group; it does not imply national or international standing, and the group's ages span 25–34 through 65+. *Casual* means rowing recreationally outside a competitive programme; it does not by itself mean inexperienced, although in this sample it largely coincided with inexperience, ten of the twelve reporting less than a year of rowing. Where later chapters describe a particular participant as a novice, that refers to their own reported experience rather than to the recruitment group.

- **Elite rowers** ($n = 5$): currently training with a club, varsity, provincial or high-performance programme, having competed at club level or above within the last 24 months. All five row two to five sessions per week; four report rowing-club membership and four report on-water rowing, with three reporting both. Four report four or more years of experience, three of them more than six. Ages spanned 25–34 through 65+; four men and one woman.

- **Casual rowers** ($n = 12$): row recreationally, not currently training with a competitive programme. Ten of the twelve report less than one year of rowing experience and ten report zero to one sessions per week; the remaining two report one to three years. Ten are students. Eleven had never used any technology-based rowing feedback system beyond the standard ergometer console. Ages were 18–24 ($n = 6$), 25–34 ($n = 5$) and 35–44 ($n = 1$); seven men and five women.

The groups separate cleanly on the objective background items. All five elite participants train at least twice weekly, against two of twelve casual participants; four of five report rowing-club membership, against none of the casual group; and four of five row on water, against two of twelve. No casual participant reports club membership at any frequency. Section 5.4 shows that the two groups are separated just as decisively by the rowing itself, which is the stronger validation of the grouping.

Exclusion criteria were a history of severe motion sickness or vestibular disorder, any medical condition making up to ten minutes of moderate-intensity rowing unsafe, and pregnancy. Two participants disclosed relevant conditions on the optional background item and were included after discussion: one casual participant reported susceptibility to motion sickness, and one elite participant reported chronic vestibular dysfunction. Both disclosures bear on the results and are reported in Section 5.9 rather than set aside.

Table 4.1: Participant characteristics by group, from the background questionnaire administered before rowing. Counts are participants. The two groups are separated on every rowing-related item and are comparable on age spread, gender balance and comfort with digital systems, so the between-group contrast in Chapter 5 is not obviously confounded by technology familiarity.

		Casual ($n = 12$)	**Elite** ($n = 5$)
Age	18–34	11	1
	35–54	1	2
	55+	0	2
Gender	Man	7	4
	Woman	5	1
Rowing experience	<1 year	10	0
	1–3 years	2	1
	4+ years	0	4
Weekly rowing	0–1 sessions	10	0
	2–5 sessions	2	5
Rowing venue	Club membership	0	4
	On-water rowing	2	4
Prior rowing technology	Used before	1	2
	None	11	3
Comfort with new systems	Comfortable or better	10	4
	Neutral or below	2	1
Weekly physical activity	Moderate or below	9	1
	High or above	3	4

4.3 Materials and Setting

The study took place in the Graphics and Experiential Media (GEM) Lab (Paramount Building, Dalhousie University; 1577 Barrington St, Halifax, NS). The setup comprises an instrumented Concept2 ergometer and a two-projector floor projection system displaying the fluid visualization described in Chapter 3, with a boat-shaped cutout masked over the machine (Section 3.3). Figure 4.2 shows what a participant sees from the seat, which is the vantage point the study's claims about peripheral attention refer to.

The console was covered. Participants saw the ergometer's own PM5 console during the unrecorded orientation period, so that they could familiarize themselves with the machine as they normally would. Before the recorded segment began, the console was covered with card taped over its face, and it remained covered for the duration of the rowing. Participants

therefore rowed with no numeric read-out of any kind: no split, no rate, no distance, no elapsed time. Figure 4.1 shows the cover in place. RowSim carries its own optional on-screen overlay reporting stroke rate, watts, pace and speed, which is enabled by default in the source and would have appeared on the projection itself; it was switched off before every recorded segment. It is the one study-relevant setting the manifest does not record, so it is stated here rather than left to be inferred from a repository default that does not describe the study.

Figure 4.1: Instrumented ergometer, photographed with the room lights on for documentation; sessions ran with them off (Section 3.3). The white sheet serving as the projection surface is visible on the floor. The card cover withholds visible PM5 metrics during the recorded segment while Bluetooth telemetry continues; the dark handle-mounted housing contains the inertial unit (Figure 3.9).

This is the design position of Section 2.3 carried into the apparatus. A thesis arguing that an under-specified peripheral display can carry an activity cannot leave a precise numeric display switched on beside it, because participants would then be reporting on a combination and nothing could be attributed to either. It also has a consequence worth stating for the results: because elapsed time was not visible, participants had no way to pace themselves against the session clock, which is part of why the duration and pausing measures in Section 5.3 record what they do.

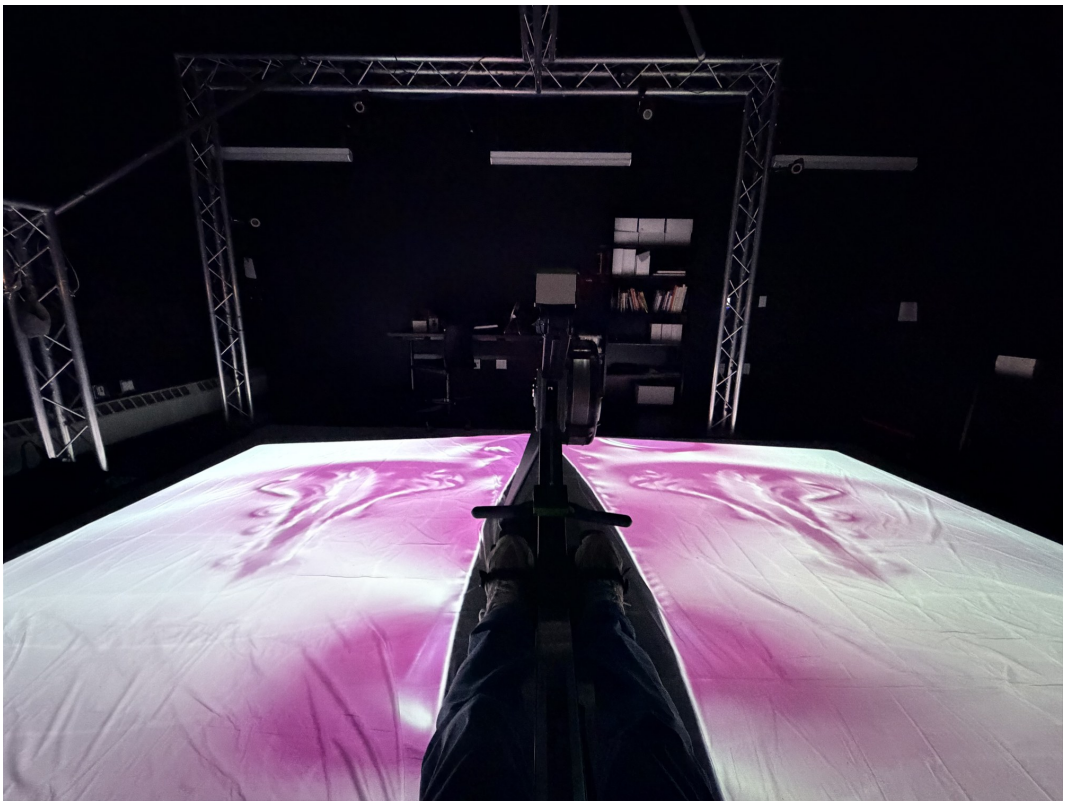


Figure 4.2: Participant-eye view from seated height. The ergometer and covered console occupy central vision, while the projection extends below and to both sides in the lower periphery. During recorded rowing it was the only visible feedback channel.

4.4 Instruments

Two questionnaires were administered after rowing: the twelve-item User Engagement Scale — Short Form (UES-SF) (89) and a custom RowSim Experience Questionnaire (REQ). Appendix B lists every item as administered together with the exact scoring.

The UES-SF supplies the conceptual and psychometric basis for the engagement measure: a multidimensional instrument, validated by its authors, covering focused attention,

perceived usability, aesthetic appeal and reward (89). One qualification belongs here rather than in the appendix. The items as administered were contextually adapted for RowSim, and three published items were replaced (Appendix B.1 sets out the mapping item by item). The published validation therefore transfers to the construction of the instrument but not to the version actually used: the adapted scale should not be treated as independently validated, and its subscale means are not directly comparable with those reported in studies using the published UES-SF. One further consequence of the adaptation deserves stating, because it affects how the overall score should be read. The administered instrument carries two focused-attention items, four perceived-usability items, and three each for aesthetic appeal and reward. The overall score is the mean of the administered items, so perceived usability contributes twice the weight of focused attention to it. That is arithmetically what it is, but it means the overall figure should be read as the mean of the items actually asked rather than as a psychometrically equivalent UES-SF total. Subscale means are unaffected. It is referred to throughout as the UES-SF for brevity, and the adaptation is documented in full in Appendix B.1. The REQ supplies what a general engagement scale cannot: rowing-specific items on rhythm, effort, interpretive load, monotony and substitution preference, together with four free-text prompts. The two are complementary rather than redundant — the UES-SF covers aesthetic appeal, which the REQ does not, and the REQ covers physical and cognitive demand and overall satisfaction, which the UES-SF does not.

Instruments considered and not used. Several validated instruments were reviewed and deliberately excluded (Table 4.2), and the reasoning is recorded here because in each case the exclusion follows from the character of the artefact rather than from convenience.

The Presence Questionnaire exclusion is the one that most directly reflects the design position. An instrument that scores an experience by how convincingly it reproduces a real environment would penalize RowSim for precisely the choices — abstraction, non-realistic colour, no depicted scene — that Section 4.1 describes as deliberate. Measuring this system against realism would have been measuring it against a goal it does not hold.

Objective data comprise the sensor and interaction logs described in Section 3.8, plus a demographic and background questionnaire administered before rowing (Appendix B.5).

Table 4.2: Instruments reviewed during design and the reason each was not administered. In every case the reason concerns fit to an abstract, continuous, physically embodied display rather than the quality of the instrument.

Instrument	Reason not administered
NASA-TLX (53, 54)	Developed for multitasking under high load. Rowing workload is captured objectively and continuously by the ergometer itself, and the REQ carries direct items on perceived physical and mental demand.
SUS (23)	Targets software and interface usability rather than continuous embodied interaction, and captures neither motivation nor experiential quality.
UEQ (71)	Overlaps conceptually with the UES-SF's aesthetic and reward dimensions; oriented to product appeal rather than to embodied feedback.
AttrakDiff 2 (57)	Strong on aesthetic quality but interface-centric, and its items presuppose a product with visual realism to evaluate.
Game- and flow-experience scales	Built around constructs such as competition, challenge and fantasy that do not apply to a repetitive, non-competitive rowing task with no goal structure (33).
Presence Questionnaire (131)	Measures spatial and sensory realism in immersive 3D environments. RowSim's visuals are deliberately abstract and non-realistic, so a realism-based presence measure would score the design's central commitment as a failure.
UES long form (88)	Thirty-one items would add roughly ten minutes after physical exertion; its Endurability and Novelty subscales concern repeated use over time and do not apply to a single session; and it duplicates REQ coverage of motivation, enjoyment and usability.

4.5 Procedure

Each session lasted approximately 50–60 minutes (Figure 4.3) and followed a written script.

1. **Introduction and consent** ($\sim$5 min). Participants reviewed and signed informed consent and selected recording options.

2. **Demographic and background questionnaire** ($\sim$3–5 min).

3. **Orientation, no simulation** ($\sim$1–2 min). Participants sat on the ergometer, adjusted the foot straps and took a few practice strokes. No research data were recorded. **No explanation of the visualization was given at any point before rowing.** Participants were not told what the display represented, which quantities drove it, or what to look for; the projection simply started. Whatever they understood of the mapping they

worked out while rowing, which is what makes the interpretive behaviour reported in Section 5.6 evidence about the display rather than recall of a briefing.

4. **RowSim rowing session** (∼10 min). Participants rowed with the ambient visualization active, under the open-ended instruction described in Section 4.1. Sensor and interaction data were logged throughout.

5. **Questionnaires**. REQ and UES-SF, answered with respect to the feedback rather than the participant's fitness.

6. **Semi-structured interview**. Nine prompts on the experience of the feedback, with follow-ups improvised in response; duration varied between participants. The full prompt list is in Appendix B.4.

7. **Debrief and honorarium**.

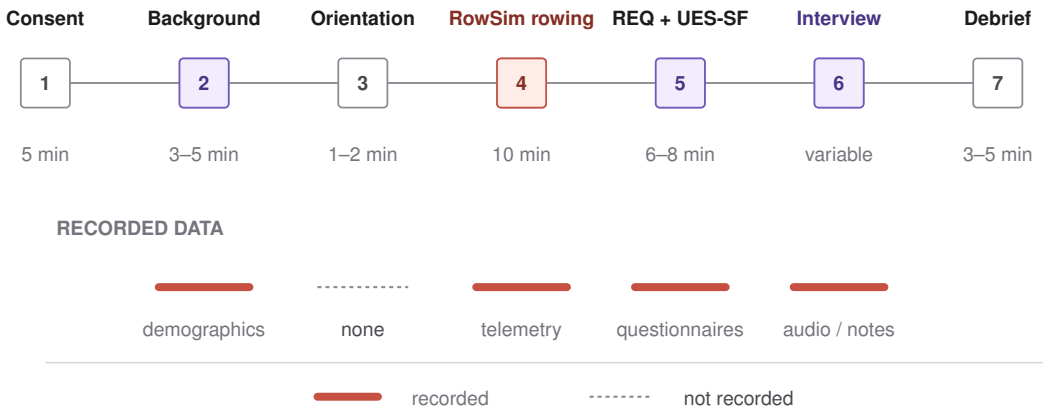


Figure 4.3: Study session flow (approximately 50–60 minutes). The lower track identifies both recording status and the data captured at each stage. The rowing segment has neither a performance target nor a continuity requirement.

4.6 Measures and Derivation

Intervals. Two intervals are distinguished, because a rate computed over one will not match a rate computed over the other. *Session duration* is the span of the logging session recorded in the manifest, from application start to session close; it includes sensor connection and the seating that precedes rowing, and is reported for transparency rather than used as a denominator. *Rowing span* is the interval from the first to the last recorded stroke, and is the exposure over which a participant interacted with the display. All rates use rowing span.

Across the full sample, session duration averaged 681 s (SD 54) and rowing span 630 s (SD 21).

Strokes. Strokes are identified from increments of the stroke counter the ergometer itself reports in the raw event stream. The derivation is validated against the ergometer's independently reported mean stroke rate, which it reproduces to within 0.15 spm. Analyses use the append-only `raw/events.jsonl` stream rather than the derived 50 Hz table, since the latter retains only one event per time bucket (Appendix A).

Pauses and bouts. An interval between consecutive strokes longer than 8 s is treated as a pause rather than a stroke cycle, and the rowing span is partitioned into bouts of continuous rowing separated by such pauses. Because pausing was permitted and never prompted, the number and total duration of pauses are reported as behavioural measures in their own right (Section 5.3), not as data to be cleaned away. Cycle-timing statistics are computed over within-bout cycles only, so that a voluntary rest does not enter the rhythm-variability estimate as an enormous stroke.

Derived quantities. Power is the mean of the ergometer's reported values while rowing. Work per stroke is mean power multiplied by rowing span and divided by stroke count; distance per stroke is session distance divided by stroke count. Both are reported because they separate force applied per stroke from the rate at which strokes are taken, which is exactly the axis on which trained and untrained rowing differ.

These measures are not independent of one another. Work per stroke is computed from mean power, and distance per stroke from session distance, so the seven force-and-output measures compared in Section 5.4 are related quantities describing one underlying difference rather than seven separate pieces of evidence. Peak drive force, drive length and the drive-to-cycle ratio come from the ergometer's own instrumentation and are not derived from the others, so the set is not wholly redundant either. The consequence for interpretation is stated where the comparison is made: the argument rests on the consistency of the pattern across related metrics, not on a count of how many crossed a threshold.

4.7 Analysis Plan

Figure 4.4 traces each research question through the instrument that measures it to the analysis that reports it.

Quantitative. Questionnaire responses are summarized by subscale and item by item. Casual and elite groups are compared using two-sided Mann–Whitney U tests, which do not require normally distributed outcomes and are appropriate for the group sizes here. Because that test operates on ranks rather than means, group comparisons are reported as median

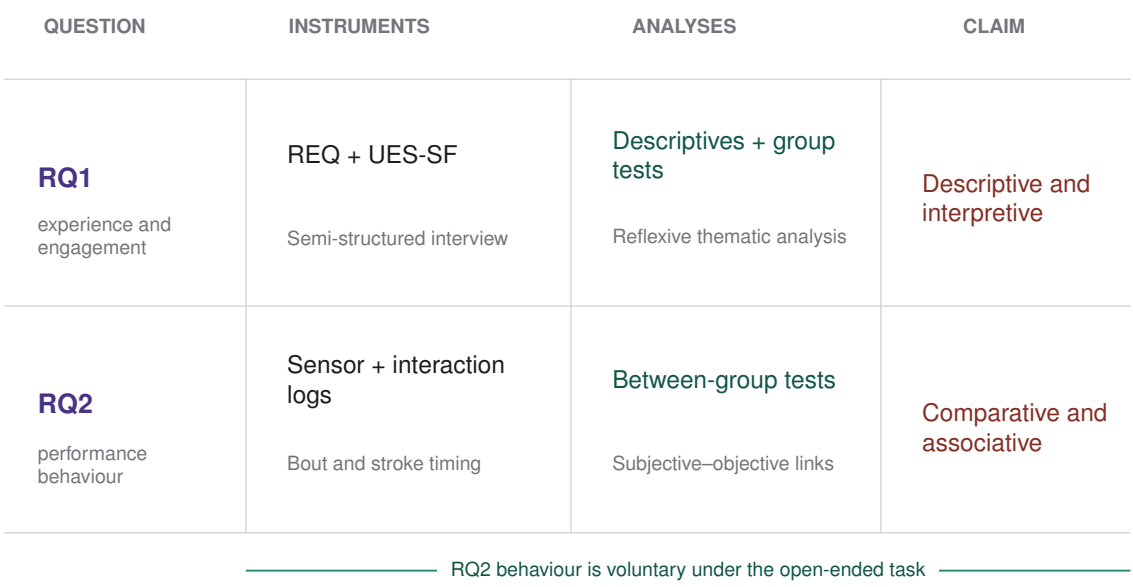


Figure 4.4: Mapping from each research question to instruments, analyses, and supported claims. Both questions combine quantitative and qualitative evidence; RQ2 behavioural measures reflect participant choice under the open-ended task.

and interquartile range with the U statistic and the rank-biserial correlation as an effect size; means and standard deviations are given separately where the distribution is of interest in its own right. Associations between subjective and objective measures use Spearman's ρ, and every correlation reported in this thesis is exploratory: none was specified in advance, none is treated as a confirmatory test, and they are read for magnitude and direction rather than for threshold crossings.

Software and p-value computation. All tests were computed with SciPy 1.15.3. `mannwhitneyu` was called with its default `method="auto"`, which enumerates the exact null distribution when the data contain no tied ranks and otherwise falls back to the normal approximation with a continuity correction. In practice this means the physical measures, which contain no ties, are reported with exact p-values, while the questionnaire measures, where several participants share a score, are reported with asymptotic ones — for example mean power returns an exact $p = .0006$ where the asymptotic approximation would give .0027, and UES-SF overall returns an asymptotic $p = .596$ where exact enumeration would give .646. The distinction changes no conclusion here, but it is recorded so the numbers can be reproduced exactly.

Scope of inference. One boundary applies to everything in Chapter 5 and is stated here rather than repeated. The study runs a single RowSim condition with no non-RowSim comparison, so it characterizes experience and behaviour *during* RowSim and examines differences and associations within the sample observed. It is not constructed to establish

that RowSim causes any change in engagement or in rowing relative to a conventional console, and no result should be read that way. One comparison is nonetheless in the data, and it should be read for what it is. The REQ asked participants directly to compare RowSim against the display they already use, and the answers were decisive: "I would prefer using the projected visuals rather than a standard rowing machine display" averaged 4.59 (SD 0.62), and its reverse, "I would choose to use a standard rowing machine instead if given a choice," averaged 1.41 (SD 0.62). That is a stated preference measured against a remembered alternative rather than a behavioural comparison against a condition anyone rowed in on the day, and it carries the weaknesses that go with self-report — novelty, the presence of the designer, no opportunity to tire of the thing. Stated preference and measured behaviour are different things, and only the first is available here. Where the literature is invoked in interpreting a pattern, it is doing so to motivate a reading, not to supply the comparison the design does not contain.

Multiple comparisons. Sixteen group comparisons are reported, in three families: force and output, timing, and engagement. Uncorrected p-values are given in the tables, because the study is exploratory and the argument does not rest on any single threshold crossing. Holm's step-down correction was nonetheless applied within each family as a check, and it does not change what can be claimed: every one of the seven force and output measures survives correction ($p_{\text{Holm}} \leq .0046$ for six of them and .027 for drive-to-cycle ratio), no timing measure approaches significance either before or after ($p_{\text{Holm}} = .70$), and no engagement measure does ($p_{\text{Holm}} = 1.00$). The pattern across measures, and the effect sizes, are what the interpretation rests on. Holm's procedure assumes nothing about dependence in the sense of remaining valid under arbitrary dependence, but it is conservative when tests are correlated, and several of these measures are derived from one another (Section 4.6); the corrected values should be read as a robustness check rather than as sixteen independent tests. Because the two groups differ so markedly in physical output, correlations are reported for the full sample and, where the distinction matters, for the casual group alone.

The group comparison carries an internal check that is worth stating in advance. The same test applied to the same two groups is used for physical measures and for engagement measures. If it separates the groups on the former and not on the latter, then the comparison is demonstrably capable of detecting large differences in this sample, and the absence of a detectable engagement difference is informative rather than vacuous. The check establishes sensitivity to differences of the size present in the physical measures; it says nothing about sensitivity to smaller ones, and Section 6.3 returns to that. Section 5.4 reports the outcome.

Qualitative. Interview and free-text material is analysed by reflexive thematic analysis, as set out below. The qualitative and quantitative strands are read against each other

rather than in sequence: several participants' accounts predict features of their own logged behaviour, and Section 5.3 reports those correspondences directly.

4.8 Qualitative Analysis

Interview and free-text material is analysed using reflexive thematic analysis (18–20). Because that approach treats the analyst as an active participant in theme construction rather than a neutral extractor, the procedure and the analyst's position are set out here rather than left implicit.

Position of the analyst. The interviews were conducted and coded by the author, who is also RowSim's designer and developer. This is a conflict worth naming: an analyst who built the artefact has an interest in reading responses favourably, and participants speaking to a system's designer have a social incentive to be generous. Three partial mitigations were used. Interview prompts were phrased neutrally and included explicit invitations to criticize ("what did you like least," "were there moments when the visuals confused you"). Themes that cut against the design position were retained and developed rather than treated as noise. Every quotation reported in Chapter 5 was checked verbatim against the source transcript by automated string matching, so quotations are not paraphrases shaped by the analyst's reading. None of this removes the conflict, and results should be read with it in view.

Data and preparation. The corpus comprises the post-session interviews and the four free-text REQ items, treated together as one dataset per participant on the grounds that participants often continued a written point verbally or vice versa. Interviews were audio-recorded where consent was given and transcribed; where recording was declined, contemporaneous written notes and written responses were used instead.

Analytic stance. Coding was primarily inductive and semantic: themes were developed from what participants said rather than from a pre-specified framework, and stayed close to explicit content rather than reaching for latent motivations. One deductive element is acknowledged — the analyst approached the corpus already holding the ambiguity-as-resource position from Chapter 2, and one theme addresses that position directly.

Process. Familiarization involved repeated reading of the full corpus before any coding. Initial codes were generated across the whole dataset, then clustered into candidate themes, reviewed against both the coded extracts and the full corpus, and revised where a candidate proved too thin or too broad to hold. Themes were then named to state what each claims rather than merely what it is about.

What the interviews can and cannot recover. The interview follows the rowing rather than accompanying it, so participants are reconstructing what they noticed rather than

reporting it as it happened, and an account of where one's attention was is not the same kind of evidence as a measurement of it. This bears on the peripherality material in particular and is taken up in Section 6.1.2.

Prevalence. Chapter 5 reports how many participants contributed to a theme. Reflexive thematic analysis does not require frequency counts and a theme's importance is not established by counting, but it is more transparent to state how widely a pattern was expressed than to leave qualifiers such as "several" undefined. Counts are of participants who raised a point unprompted anywhere in their corpus, established by reading each participant's full responses rather than by keyword search; where a count is given, the participant identifiers are listed so it can be checked.

Consent, withdrawal, and data governance. Participants gave written informed consent before any study activity and could withdraw at any point without penalty. Identifiable data could be withdrawn up to 14 days after a session; after that, analysis proceeds on coded, de-identified data. De-identified data and signed consent forms are retained for five years. Raw sensor logs, recordings and questionnaire responses are held on an encrypted Dalhousie-managed workstation.

Chapter 5

Results

5.1 Participants and Sessions

Seventeen participants completed the study: twelve casual rowers and five elite rowers (Section 4.2). Sessions averaged 681 s of logging (SD 54) containing 630 s of rowing (SD 21), during which participants took 237 strokes on average (SD 52) at 22.8 strokes per minute (SD 4.1), covering 1,713 m (SD 607) at a mean 61.6 W (SD 52.1).

The spread in that last figure is the first thing worth noting. Mean power ranged from 8.2 W to 165.7 W — a twentyfold range within a task that every participant performed under the same instruction. That range is not noise. It is what an open-ended task produces when rowers of very different backgrounds are told to row naturally and left alone.

5.2 Engagement and Experience

5.2.1 Questionnaire outcomes

Self-reported engagement was high and consistent. Across the full sample the adapted UES-SF (Section 4.4; nine of the twelve published items, three replaced) returned an overall mean of 4.37 (SD 0.43) on a five-point scale, and the REQ composite 4.54 (SD 0.32). Every participant's overall engagement score fell between 3.42 and 5.00; no participant returned a score below the scale midpoint on any subscale.

The REQ items locate that engagement precisely; Table 5.2 gives every item by group. Interpretive load was low: "I found the experience confusing" averaged 1.35 (SD 0.61) and "I felt that the experience required a lot of mental effort" 1.18 (SD 0.39). Disruption of rowing was near-absent: "I had difficulty maintaining my rowing rhythm because of the projected visuals" averaged 1.06 (SD 0.24), the lowest-variance item in the instrument, with sixteen of seventeen participants giving the minimum possible response; "I felt distracted by the projected visuals" averaged 1.29 (SD 0.47). Endorsement was near-unanimous: "I would recommend this experience to others" averaged 4.82 (SD 0.39) and "Overall, I enjoyed the

Table 5.1: Subscale and overall scores on the adapted UES-SF. Because three published items were replaced, these values index engagement within this study and are not directly comparable with published UES-SF norms (Section 4.4). Higher indicates greater engagement; perceived-usability items are reverse-scored so that higher means less friction. Focused attention is the only dimension that does not sit near the scale ceiling, and Section 5.2.2 shows why.

Subscale	**All** ($n = 17$)	**Casual** ($n = 12$)	**Elite** ($n = 5$)
Focused attention	3.82 (0.79)	3.88 (0.88)	3.70 (0.57)
Perceived usability	4.62 (0.42)	4.62 (0.39)	4.60 (0.52)
Aesthetic appeal	4.31 (0.58)	4.47 (0.41)	3.93 (0.80)
Reward	4.45 (0.50)	4.50 (0.46)	4.33 (0.62)
Overall	**4.37 (0.43)**	**4.43 (0.39)**	**4.22 (0.53)**
REQ composite	4.54 (0.32)	4.56 (0.31)	4.47 (0.39)

experience" 4.76 (SD 0.44). Asked to compare against their existing option, participants preferred the projection: "I would prefer using the projected visuals rather than a standard rowing machine display" averaged 4.59 (SD 0.62), against 1.41 (SD 0.62) for preferring a standard machine.

5.2.2 Absorption is not the same as losing oneself

Focused attention is the one subscale that does not approach the ceiling, and the reason is a split between its two items rather than uniform moderation. "I was absorbed in the activity" averaged 4.41 (SD 0.62); "I lost myself in this experience" averaged 3.24 (SD 1.44) — the largest standard deviation of any item in either instrument. Four participants scored these two items three or more points apart.

Part of that variance may have a plain cause, and it should be reported before any interpretive one. This was an unprompted in-session observation rather than a coded response, so no participant count is available for it: while completing the questionnaire, participants asked aloud whether "lost" was meant positively or negatively — whether losing oneself in the experience was something the scale wanted them to affirm or to deny. The item is not ambiguous about its construct, which is absorption; it is ambiguous about its valence, because "lost" carries an ordinary negative sense (disoriented, adrift) alongside the technical positive one the scale intends. Participants who read it the first way and participants who read it the second way would answer the same felt state at opposite ends of the scale, which is one plausible contributor to an SD of 1.44 in an instrument where no other item exceeds 1.17. Section 6.4 takes this up as a problem about measuring experience rather than a defect in this administration.

Table 5.2: Every REQ item, mean (SD), by group. Positively worded items are grouped first and negatively worded ones second, so the two blocks can be read at a glance: endorsement sits near the top of the scale and every burden item sits near the floor, for both groups. The physical-effort item is listed separately because it is excluded from the composite — high physical effort is expected in a rowing task and is neither favourable nor unfavourable. Item wording as administered is in Appendix B.2.

Item	**All** ($n = 17$)	**Casual** ($n = 12$)	**Elite** ($n = 5$)
Positively worded — higher is more favourable			
Easy to understand	4.53 (0.72)	4.58 (0.67)	4.40 (0.89)
Visuals matched my actions	4.18 (0.73)	4.33 (0.65)	3.80 (0.84)
Learned quickly	4.29 (0.92)	4.17 (0.94)	4.60 (0.89)
Made me want to row more	4.18 (0.81)	4.25 (0.75)	4.00 (1.00)
Felt absorbed	4.00 (1.12)	3.92 (1.31)	4.20 (0.45)
Kept my attention	4.47 (0.72)	4.67 (0.49)	4.00 (1.00)
Reduced monotony	4.35 (0.86)	4.42 (0.90)	4.20 (0.84)
Enjoyed overall	4.76 (0.44)	4.75 (0.45)	4.80 (0.45)
Prefer to a standard display	4.59 (0.62)	4.75 (0.45)	4.20 (0.84)
Would recommend	4.82 (0.39)	4.92 (0.29)	4.60 (0.55)
Useful for training or rehab	4.53 (0.72)	4.67 (0.65)	4.20 (0.84)
Negatively worded — lower is more favourable			
Found it confusing	1.35 (0.61)	1.42 (0.67)	1.20 (0.45)
Unsure what visuals meant	1.94 (0.90)	2.00 (0.85)	1.80 (1.10)
Felt distracted	1.29 (0.47)	1.25 (0.45)	1.40 (0.55)
Boring or disengaging	1.18 (0.39)	1.17 (0.39)	1.20 (0.45)
Difficulty keeping rhythm	1.06 (0.24)	1.08 (0.29)	1.00 (0.00)
Required mental effort	1.18 (0.39)	1.25 (0.45)	1.00 (0.00)
Made rowing more tiring	1.12 (0.33)	1.17 (0.39)	1.00 (0.00)
Would choose standard machine	1.41 (0.62)	1.42 (0.67)	1.40 (0.55)
Excluded from the composite			
Required physical effort	1.88 (1.17)	2.08 (1.24)	1.40 (0.89)

The interviews explain the remaining divergence, and it is not measurement error. Participants distinguished being absorbed *in* the activity from losing themselves. Some described the latter directly: P4 reported that "The 10 minutes time period doesn't feel for 10 minutes… felt like 2 or 3 minutes"; P15 that the session "felt like I am somewhere else." Others described sustained, deliberate, fully self-aware attention. P8 was explicit that he was "purely trying to figure out how what I did affected the visuals." P10, an elite rower, reported that "I was trying to trip it up at times, and it didn't get confused as much as I thought."

Both groups rated the system highly overall; neither was disengaged. But only the first is what an absorption item is built to detect. Section 5.3 shows that these two modes also differ behaviourally.

5.3 Rowing Behaviour Under an Open Task

Because participants were given no target and were free to stop (Section 4.1), how long they rowed and whether they paused are choices rather than compliance, and can be read as behaviour.

Thirteen of seventeen participants never stopped rowing at all. Across a nominally ten-minute session with an explicit invitation to rest, thirteen produced an unbroken stroke sequence from first stroke to last, with no gap exceeding eight seconds. Four paused: P8 seventeen times for 175 s in total, P5 six times for 64 s, P15 six times for 62 s, and P10 twice for 39 s.

That distribution is itself the finding, and it is corroborated from two directions. Figure 5.1 shows every session as a timeline, which makes the shape of each participant's choice visible in a way the counts cannot.

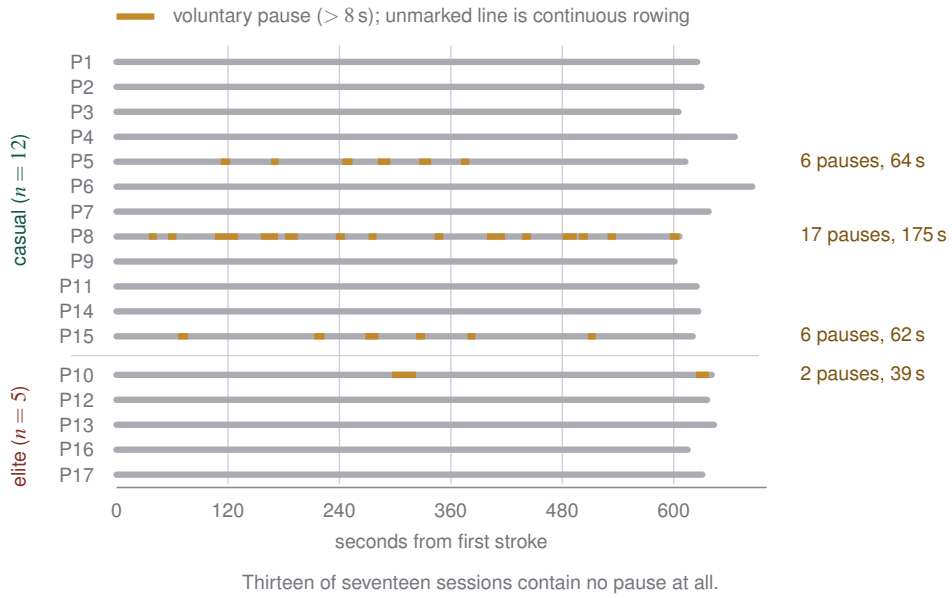


Figure 5.1: Sessions from first to last stroke; amber marks voluntary pauses longer than eight seconds. Thirteen participants rowed continuously. Casual pauses were brief and scattered, whereas P10's longer breaks align with self-imposed rowing pieces.

First, participants who did not pause said so, unprompted, in terms of the display. P7 reported that the visuals "did motivate me to keep up with my pace" and that "I didn't want

to take any breaks or whatever because it was quite engaging" — and P7's log contains no pause. P11 was more explicit about the counterfactual: "at a gym environment, I'm not sure if I was going to continue rowing this much," adding "Here, without taking breaks. I continued to do that because it felt good." P11 likewise has no recorded pause. Two participants predicted their own logged behaviour, in a session where neither could see the log and neither was asked about breaks.

Second, the three casual participants who did pause are the three whose interviews contain the corpus's clearest reservations. P5 volunteered that "the same patterns again and again, definitely overwhelmed after some time because I did not want to look at it after 5, 6, 7 minutes." P15 reported that it "took me a while to understand how it's reacting." P8 reported mild motion sickness on sustained viewing (Section 5.9). Among the twelve casual participants, total paused time correlates with overall engagement at $\rho = -0.567$ ($p = .054$) — the strongest association between any behavioural measure and any self-report measure in the dataset.

The fourth pauser breaks the pattern in a way that is more informative than the pattern itself. P10 is an elite rower, returned one of the highest engagement scores in the study (4.75), and produced the second-highest power output (144.2 W). His two pauses were 26 s and 14 s, and his interview says what they were for: "I did a couple pieces in there where, I either put on more pressure or I put on more rate and pressure," and "I did a start in there where I was trying to let the hull run down." Given an open task, the most experienced participant imposed a training structure on it — discrete pieces separated by rest — which is what a competitive rower does with an unstructured ten minutes. This also accounts for his stroke-timing variability (coefficient of variation, CV, of 24.6%, the second-highest in the study), which reflects deliberate rate changes between pieces rather than an unstable rhythm.

The same behaviour therefore means opposite things in the two groups, and only the interviews distinguish them: for three casual participants pausing indexed waning engagement, and for one elite participant it indexed expert self-direction. A purely quantitative reading would have merged them.

5.4 Casual and Elite Rowers Compared

The two groups differ enormously in what they did with the ergometer, and on no engagement measure is a difference detectable.

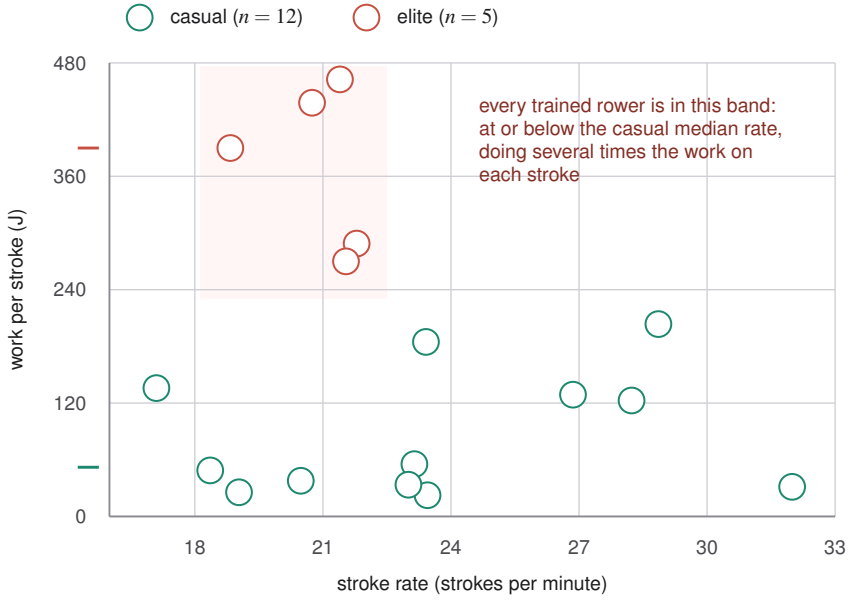
The physical separation is large, and in the expected direction. Comparing group means, elite rowers applied 2.6 times the peak drive force, moved the handle 1.7 times as far, produced 3.7 times the mean power, and did 4.3 times the work per stroke — the last

Table 5.3: Every measure compared between groups. Engagement rows use the adapted UES-SF and the REQ (Section 4.4); neither is directly comparable with published norms. Measures are given as median [interquartile range] with the Mann–Whitney U statistic, the two-sided exact p, and the rank-biserial correlation r_{rb} as an effect size. r_{rb} is the proportion of casual–elite pairs favouring one group minus the proportion favouring the other, so 1.00 means every elite value exceeds every casual value and 0 means the groups are interleaved. The force and output block separates completely — three measures at $r_{rb} = 1.00$ — while no engagement measure exceeds 0.42. Holm-corrected p-values within each family are given in Section 4.7.

	Median [IQR]				
Measure	**Casual** ($n = 12$)	**Elite** ($n = 5$)	U	p	r_{rb}
Force and output					
Peak drive force (N)	29.4 [22.5–73.0]	124.4 [114.5–128.6]	1	.0006	0.97
Drive length (m)	0.63 [0.57–0.75]	1.21 [1.11–1.21]	0	.0003	1.00
Mean power (W)	19.1 [12.7–57.9]	123.0 [105.3–144.2]	1	.0006	0.97
Work per stroke (J)	52.1 [33.1–130.6]	389.9 [288.8–437.8]	0	.0003	1.00
Distance per stroke (m)	5.73 [4.84–7.25]	10.67 [9.69–12.25]	0	.0003	1.00
Session distance (m)	1263 [1036–1799]	2251 [2249–2379]	3	.0023	0.90
Drive-to-cycle ratio	0.32 [0.30–0.35]	0.26 [0.26–0.27]	9	.027	0.70
Timing					
Stroke rate (spm)	23.3 [20.1–27.2]	21.4 [20.7–21.5]	18	.234	0.40
Cycle CV (%)	13.8 [10.1–19.9]	8.6 [8.3–8.9]	19	.279	0.37
Strokes	244 [210–278]	226 [211–232]	21	.370	0.30
Engagement					
UES-SF overall	4.38 [4.12–4.67]	4.33 [4.00–4.58]	24	.596	0.18
UES-SF focused attention	4.00 [3.00–4.62]	3.50 [3.50–4.00]	26	.747	0.12
UES-SF perceived usability	4.75 [4.50–4.81]	4.75 [4.50–5.00]	29	.956	0.03
UES-SF aesthetic appeal	4.50 [4.00–4.75]	4.00 [3.33–4.33]	18	.195	0.42
UES-SF reward	4.33 [4.00–5.00]	4.00 [4.00–5.00]	23	.466	0.23
REQ composite	4.61 [4.41–4.71]	4.68 [4.37–4.68]	28	.915	0.05

of these at or below the casual group's stroke rate rather than above it, which Figure 5.2 shows as a distinct region of the rate–work plane rather than a shift along it. All seven output measures show nominally significant group differences, and all seven survive Holm correction within the family (Figure 5.3 ranks every measure tested by how far apart the groups sit) — six at $p_{Holm} \leq .0046$ and the drive-to-cycle ratio at .027 (Section 4.7). These are not seven independent results: work per stroke is derived from power and distance per stroke from session distance (Section 4.6), so what the block establishes is one difference in physical output showing up consistently wherever it is measured, rather than seven separate confirmations of it. Three of the seven separate the groups *completely*: not one casual participant reaches the lowest elite value for drive length, work per stroke or distance per

stroke ($r_{rb} = 1.00$ in Table 5.3). The other four overlap by a single participant — P6 exceeds the lowest elite value for mean power (98.4 against 97.3 W) and for session distance, P1 does so for peak drive force, and P1 and P8 fall inside the elite range on drive-to-cycle ratio — so the separation is strong rather than total. The pattern is the textbook signature of trained rowing: length and force per stroke rather than stroke frequency (67, 110). Elite rowers in fact rowed *slower* in rate terms — a median 21.4 against 23.3 spm — while covering a median 10.67 m per stroke against 5.73 m. Their drive-to-cycle ratio was correspondingly lower (median 0.26 against 0.32, $p = .027$), meaning a proportionally longer recovery, which is what rowing coaching calls good ratio. That group assignment made at recruitment is reproduced this cleanly by the ergometer's own telemetry is independent confirmation that the two groups are what they are labelled.



Each point is one participant's session mean. Rules in the left margin mark the group medians on the vertical axis.

Figure 5.2: Stroke rate versus work per stroke. Elite rowers occupy a distinct high-work region while rowing at or below the casual median rate, producing a larger group difference in work per stroke (4.3×) than in rate.

Engagement does not separate the groups. Overall UES-SF differs by 0.21 scale points and the REQ composite by 0.09. No engagement measure approaches significance; the closest is aesthetic appeal at $p = .195$, and Section 5.8 suggests why that is the one that moves.

The juxtaposition is the point, and its force comes from the fact that both halves of the table are the same test on the same seventeen people. A null result invites the objection that five elite participants cannot detect anything. That objection is answered within the table: the identical comparison, on the identical participants, detected differences at $p \leq .0006$ on

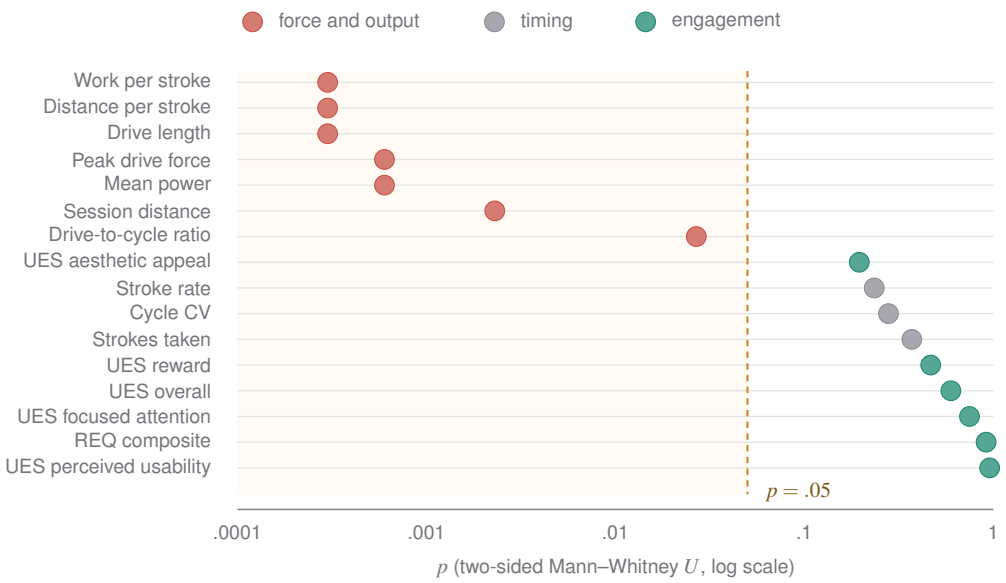


Figure 5.3: Casual–elite comparisons from Table 5.3, ordered by exact p-value. All seven physical-output measures separate the groups, whereas none of the six engagement measures does. Non-detection does not establish equivalence.

five measures and at $p \leq .027$ on all seven. The comparison had ample sensitivity to large group differences and found them — on how the rowing was done. It found none on how the display was experienced.

What follows is bounded but real. Overall engagement did not differ detectably between the casual and elite groups despite large differences in physical output. Figure 5.4 puts the two side by side: physical output separates cleanly, engagement does not. A participant producing 8.2 W and a participant producing 165.7 W returned overall engagement scores of 4.00 and 4.33 respectively, and the twenty-fold range of physical output between the extremes of this sample is spanned by 1.58 scale points of engagement.

5.4.1 Where the groups did differ

Two item-level patterns are worth recording even though the subscales do not separate, because together they point in a coherent direction.

The elite group found the display easier, not harder. All five gave the minimum response to "required a lot of mental effort" (1.00, SD 0.00), to "I had difficulty maintaining my rowing rhythm" (1.00, SD 0.00), and to "The projected visuals made rowing feel more tiring than usual" (1.00, SD 0.00), against casual means of 1.25, 1.08 and 1.17. They also learned it faster: "I was able to quickly learn how to affect the projected visuals" returned 4.60

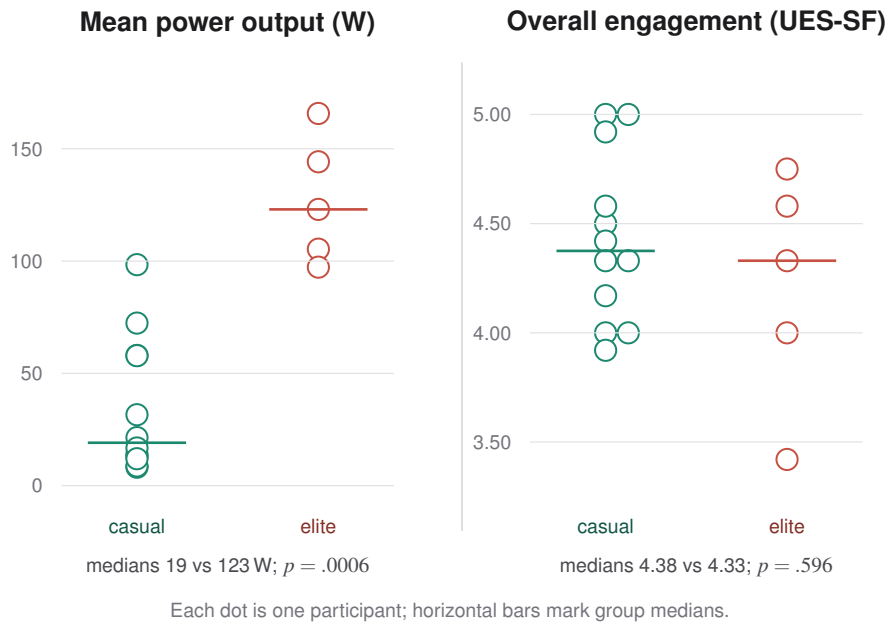


Figure 5.4: Physical output and engagement for the same seventeen participants. Output distributions barely overlap and their medians differ sixfold; engagement distributions overlap and their medians differ by 0.05 scale points.

(SD 0.89) against 4.17 (SD 0.94). That is not a foregone conclusion for an abstract display shown to people with strong prior expectations of what rowing feedback looks like.

Yet the elite group was more conservative about substitution. "I would prefer using the projected visuals rather than a standard rowing machine display" returned 4.20 (SD 0.84) from elite participants against 4.75 (SD 0.45) from casual ones, and "The experience made me want to row more" 4.00 against 4.25. Enjoyment was if anything slightly higher — "Overall, I enjoyed the experience" returned 4.80 against 4.75 — so the reservation is about replacing an instrument they depend on rather than about the experience. P10 said as much directly, describing what he would want added rather than removed: "bring all of that data that's coming from the PM 5 and lay it out," with "the ability to show or hide whatever you want to see."

Table 5.4: Every participant's session, so that the group statistics above can be checked against the individual records they summarize. C denotes casual, E elite. Span is the interval from first to last stroke; CV is the coefficient of variation of within-bout cycle time; Pau is the number of pauses longer than eight seconds. Participant labels in bold mark the rows the text returns to: P6 (highest casual output with the lowest casual variability), P8 (seventeen pauses), P10 (self-imposed pieces), P16 (lowest variability and lowest engagement in the study) and P17 (highest output).

P	Grp	Span (s)	Strokes	spm	CV (%)	Power (W)	Dist (m)	Pau	UES	REQ
P1	C	625	245	23.4	17.7	72	1903	0	4.50	4.68
P2	C	630	216	20.5	19.2	13	1056	0	4.42	4.63
P3	C	605	193	19.0	14.6	8	903	0	4.00	4.16
P4	C	666	262	23.5	11.4	9	2103	0	5.00	4.95
P5	C	613	175	18.4	22.1	14	978	6	4.33	4.68
P6	C	685	331	28.9	6.3	98	2341	0	4.17	4.58
P7	C	638	301	28.2	4.6	58	1765	0	4.33	4.47
P8	C	606	141	17.1	38.5	32	1231	17	4.00	4.00
P9	C	601	270	26.9	6.2	58	1694	0	4.58	4.58
P11	C	625	242	23.1	12.4	21	1295	0	5.00	5.00
P14	C	626	335	32.0	13.0	17	1156	0	4.92	4.79
P15	C	620	221	23.0	37.0	12	902	6	3.92	4.21
P10	E	641	211	20.7	24.6	144	2251	2	4.75	4.68
P12	E	636	232	21.8	8.9	105	2139	0	4.58	4.79
P13	E	643	232	21.5	8.3	97	2249	0	4.00	4.37
P16	E	615	194	18.8	3.3	123	2379	0	3.42	3.84
P17	E	631	226	21.4	8.6	166	2768	0	4.33	4.68

5.5 Associations Between Experience and Behaviour

Table 5.4 lists every participant's session in full. Spearman correlations between the self-report measures and the behavioural measures are given in Table 5.5.

Three patterns stand out.

No association between engagement and how hard participants rowed is detectable. Both instruments return correlations indistinguishable from zero with mean power ($\rho = -0.07$ and -0.01), consistent with the group comparison in Section 5.4.

Engagement shows an exploratory association with whether participants kept going, within the casual group only, that falls just short of the conventional threshold. Total paused time carries the strongest association with engagement in the dataset at $\rho = -0.57$ ($p = .054$) among casual participants, and weakens to -0.27 across the full sample precisely because elite pausing is self-directed rather than disengaged (Section 5.3). That the association

Table 5.5: Spearman correlations between self-reported engagement and behavioural measures. The full-sample column includes all seventeen participants; the casual column excludes the elite group, whose pausing has a different meaning (Section 5.3).

Self-report	Behaviour	**All** ($n = 17$) ρ (p)	**Casual** ($n = 12$) ρ (p)
UES-SF overall	mean power	-0.07 (.80)	$+0.07$ (.84)
UES-SF overall	session distance	$+0.04$ (.89)	$+0.46$ (.14)
UES-SF overall	stroke rate	$+0.47$ (.06)	—
UES-SF overall	cycle CV	$+0.04$ (.88)	-0.48 (.12)
UES-SF overall	total paused time	-0.27 (.30)	$\mathbf{-0.57}$ (**.05**)
UES-SF focused attn.	mean power	-0.34 (.18)	—
UES-SF focused attn.	work per stroke	-0.43 (.09)	—
REQ composite	mean power	-0.01 (.96)	-0.02 (.96)
REQ composite	total paused time	-0.23 (.37)	-0.41 (.19)

is diluted rather than strengthened by adding participants is consistent with the interview evidence that pausing served different purposes in the two groups; the interviews, not the coefficient, are what distinguish them.

Focused attention runs in the opposite direction to output. Absorption correlates negatively with work per stroke ($\rho = -0.43$, $p = .085$) and with mean power ($\rho = -0.34$). The direction matches the interview accounts in which absorption is described as a release from effort rather than an accompaniment to it, and Chapter 6 takes up what that implies.

One full-sample coefficient should be read with care. Engagement correlates with stroke rate at $\rho = +0.47$ ($p = .057$), the largest full-sample association in the table. This may reflect between-group structure rather than an association within either group: elite rowers row at lower rates by training, and one elite participant returned the lowest engagement score in the study, so the coefficient partly re-expresses the group difference in rate. Within the casual group alone no comparable association appears.

How the responses were distributed

Means conceal the structure of these responses. Figure 5.5 gives every Likert item from both instruments with its full distribution, ordered by standard deviation, and two things become visible that a table of means cannot show.

At the tight end, agreement is near-unanimous and it is concentrated on exactly the items that bear on non-disruption. Sixteen of seventeen participants gave the lowest response on the scale — strongest possible disagreement — to having had difficulty maintaining rhythm

because of the display; fifteen did the same for the display making rowing feel more tiring, and fourteen for its requiring a lot of mental effort. These are the most consistent responses in the study, and all of them say the same thing: the display did not get in the way.

At the other end sits one item alone. *"I lost myself in this experience"* returns a standard deviation of 1.44 — the largest of any of the thirty-two items, and slightly more than double that of any other item on its own scale, where the next widest is 0.71. Its responses spread across the entire scale rather than clustering anywhere. What makes this specific rather than a general looseness at the top of the ordering is what the two next-widest items are: physical effort required (1.17) and absorption while rowing (1.12). Both ask participants about their own state during exercise, where genuine variation between a novice and a trained rower is expected and unremarkable. No other item about the display itself comes close. Section 6.4 argues that the spread on *lost* is therefore a property of the item rather than noise in the sample, and Figure 5.5 is the ground for treating it that way: an item that means different things to different respondents should look exactly like this, and nothing else asked about the system does.

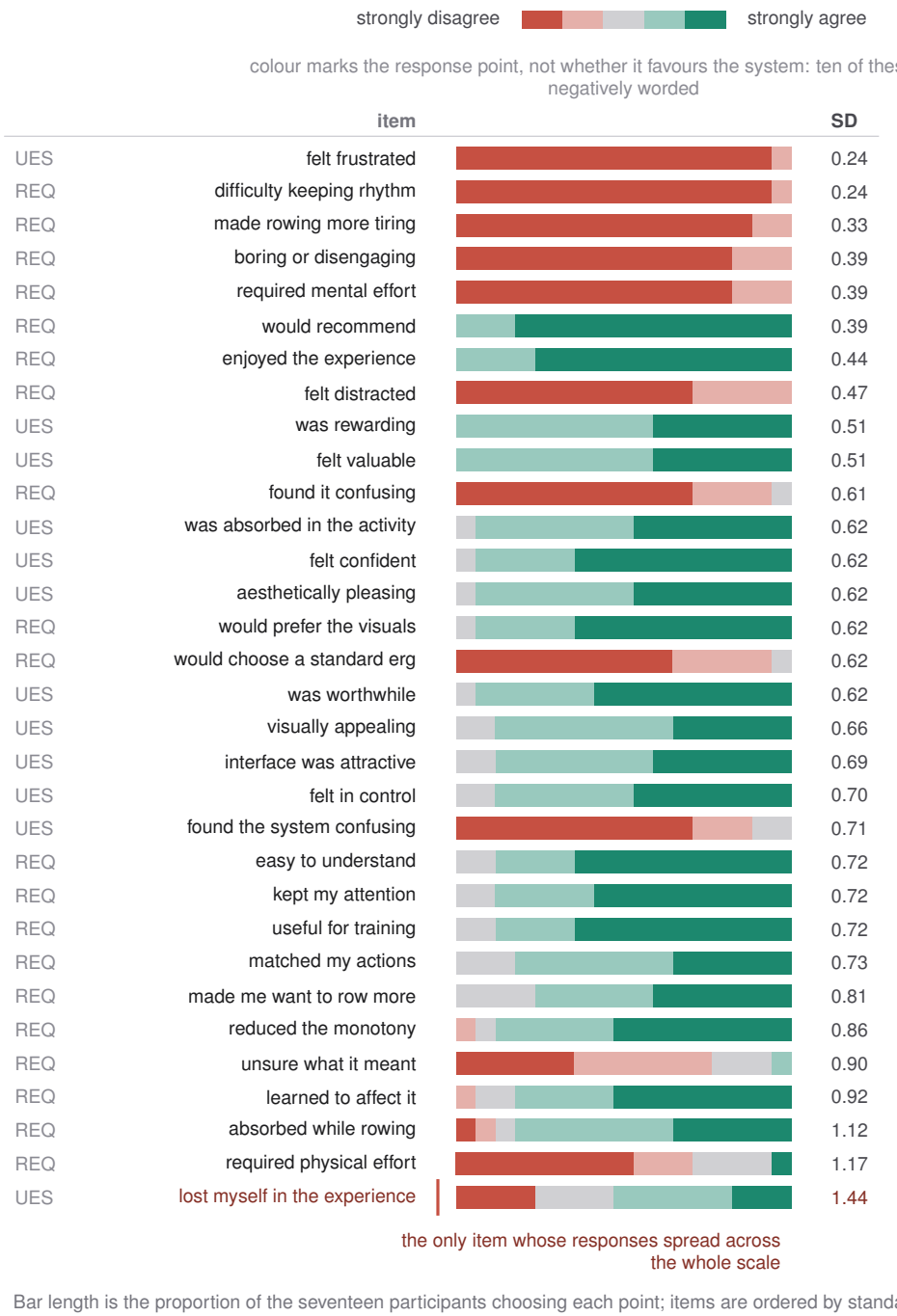


Figure 5.5: Response distributions for all 32 Likert items, ordered by standard deviation. Participants agreed most strongly on non-interference; the UES-SF "lose myself" absorption item spans the full scale and has the study's largest dispersion.

5.6 Qualitative Findings

Reflexive thematic analysis of the seventeen post-session interviews and the free-text REQ responses produced six themes, summarized in Figure 5.6 with the design commitment each one bears on; the procedure and the analyst's position are set out in Section 4.8. Counts follow the convention set out in Section 4.8. Each theme names the participants who contributed to it, so that a reader can see which claims rest on convergence across the sample and which rest on a small number of articulate accounts.

Theme	Bearing on the thesis's design position
1 **Coupling drove engagement**	supports the real-time mapping in Section 3.9
2 **Ambiguity invited exploration**	directly supports ambiguity-as-resource (Section 2.3)
3 **Attention left the numbers**	experts and novices converge on this independently
4 **Globally legible, locally opaque**	qualifies it: some features were never decoded
5 **Stopping was not commandable**	a specific, actionable gap in the interaction model
6 **Participants predicted novelty decay**	the sharpest challenge to the design position

Badge colour marks direction: teal supports a design commitment, purple qualifies it, coral challenges it.

Figure 5.6: The six themes and what each does to the thesis's design position. Three support it, two qualify it, and one challenges it directly.

Theme 1: Coupling was the source of engagement

Participants located the system's value in the immediate, legible correspondence between what they did and what the floor showed. All seventeen described this correspondence in some form, making it the most widely expressed theme in the corpus. P1: "For every stroke I was able to see how much water or the fluid I'm able to displace." P2: "I could see exactly the moment when I was pushing back." P6: "the stronger I'd pull the row, the faster I felt

I was going." P14 described using it to calibrate effort: "how much force I need to put to change my rowing speed." P13, an elite rower, put it in terms of control: "it was synced to how I was moving, so I like that I could affect how the visuals were going." P17, an elite rower, was precise about it: "The timing of the swirl patterns was quite accurate with the drive on the erg," and "when I changed my stroke rate. It immediately, swirl patterns changed exactly right." The theme maps onto the REQ item "the projected visuals matched my actions in a meaningful way" ($M = 4.18$).

Two participants at opposite ends of the expertise range described the specific perceptual effect the system was built to produce. P4, a casual rower, wrote that "it gives me the sense of i'm moving." P17, the most experienced rower in the sample, described the same thing at length, and described it as a change in what the session was like rather than as a feature: "I enjoyed the sense that I was moving through space. It did simulate that fairly well, as opposed to just being in a gym full of a bunch of other people," and "it was nice to have that sense, and it improved the erg-ing experience for me." Section 6.1.6 takes up the one respect in which that sense fell short. P10, who has rowed for more than thirty years and uses indoor tanks, placed it against the alternatives available to him: "The visuals did provide a simulated on-water experience, more so than other systems on the current market."

For P13 the coupling was specifically rhythmic rather than informational: "I like the rhythm it created. Rowing on the machine is very boring," and "Having that visual enhances the feeling of being in a good rhythm." The display was valued not for what it reported but for what it did to the felt quality of a repetitive action.

One further observation belongs with this theme, because it is easy to lose behind the subscale means. Asked at the end whether there was anything else they wanted to say, participants volunteered plain verdicts on the experience rather than on the interface: P8, for whom this was a first exposure to rowing of any kind, "I thought it was very interesting and fun to use this tool"; P9, "I had quite fun using the rowing machine"; P15, "It was fun. Yeah, fun experience"; P5, that rather than thinking about exercising, "I actually been experiencing something pleasant"; P6, simply, "I like the experience." P10, whose thirty years in the sport make him the least likely person in the sample to be impressed by novelty, offered an assessment of the work rather than of the session: "It's a great project. I don't think I've ever heard of anything like this. I haven't seen anything like it online" — an assessment he qualified by noting how closely he follows the sport. The subscale means do not carry this. Enjoyment was not the construct the study set out to measure, but it is the register most participants reached for once the questions stopped constraining them, and a report that gave only the numbers would describe a quieter session than the one that happened.

Two accounts go past enjoyment to the question the thesis is actually about, which is whether anyone would use such a thing. P8, rowing for the first time, located the difference in what the machine alone affords: what stood out was "just how fun it was to play with and experiment with. I feel like on a normal rowing machine, I wouldn't have gotten that experience," and "it did feel more immersive than just the machine itself." P7 reported a behavioural consequence rather than an attitude — "I didn't want to take any breaks or whatever because it was quite engaging" — which is engagement expressed as something not done. P12 approached it from thirty-five years of competitive rowing and named the problem the thesis was built around before being asked about it: "Most of us love to row in the water, but we don't really like the erging experience, right?" He then said what would follow: "if I was using this regularly for my erg winter workouts, which are kind of long… this would enhance that experience, make the time go faster, be much more pleasant" — forty minutes or more, in his description. A single session cannot establish that it would; what it establishes is that the person best placed to know the alternative thought it might.

Theme 2: Ambiguity produced exploration rather than confusion

Rather than seeking a legend, participants probed the mapping. Ten of seventeen described deliberate experimentation (P3, P4, P6, P8, P9, P10, P13, P14, P15, P17).

The clearest case is P6, who ran a systematic comparison: "I was pulling my elbows up, pulling my elbows down, and turning my wrists up… curious to see if it would change anything." P6 reached a definite conclusion — that the elbows-down technique produced the most speed — and validated it against the display: "They matched the feedback." This is notable because P6 also returned the second-highest power output among casual participants (98.4 W) and the lowest stroke-timing variability in the entire study (CV 6.3%). The participant who explored the mapping most systematically also rowed the most consistently.

Others probed differently. P3 varied hand height "just to see whether that bubble thing changes." P8 "felt more interested to try different things with the simulation." P9 "was trying to sync up with it." P15 said the display "helped me gauge which actions would lead to what consequences." P13 described the same behaviour as curiosity rather than analysis: "it's more that you have a curiosity with, 'Oh, how can I change the visuals? Let me change how I'm rowing just to see.' "

P10's version is the most demanding, and the most useful. He set out to break the system: "I was trying to trip it up at times, and it didn't get confused as much as I thought." He then probed the specific boundary he cared about, reporting that puddle size did not always scale with pressure during rate changes while stabilizing once he settled: "if I was, like, in

a steady state where I was rowing, at even pressure, and at even rate, it seemed to be, it seemed to level out." An expert user stress-testing an ambiguous mapping produced the most precise account of its limits in the corpus.

P16 is the instructive exception, and the only participant who reports not having probed at all: "for this session, actually, I think I rowed at pretty steady speed, so I didn't really see much of how my speed affects the visuals, whether the projected visual moves faster or slower because of my speed." His session bears this out — a cycle-time CV of 3.3%, the lowest in the study, from a rower who says he "kept to my rhythm, and so I wasn't affected by any external factor." He also returned the lowest overall engagement score in the study (3.42). A participant who did not vary his rowing did not discover the coupling, and did not find the display as engaging as those who did. With one case this is a correspondence rather than a mechanism, but it is the correspondence the calibration account predicts.

Not every interpretation converged on the designed one. P15 constructed a physically reasoned explanation for an effect the system does not implement: "when I moved the handle, it changed the color on that side. It went from pink to white," reasoned to be because "when the water's angle changes in real life, the color also changes a little bit." The account is wrong about the mechanism and coherent as physics. P16 did something similar with elements he could not place: "I see the two things on the two sides of the machines, and I'm not sure if they actually mean anything… I suppose they're supposed to represent some kind of fish, but I'm not entirely sure." Both are meaning-making of the kind the ambiguity literature describes, rather than the convergent calibration that most participants performed.

Theme 3: Attention moved away from the numbers

Participants described the display displacing the numeric console, and did so from both ends of the expertise range.

Among casual participants: P1 rowed "instead of looking into the ergometer to track my speed"; P9 "instead of just looking at random numbers." P9 went further and named the mechanism: "when I was usually looking at a metric, I'm sort of optimizing for the metric in particular. But when I do not have a metric, I'm just sort of having fun with it to some extent."

P10, with more than thirty years of rowing behind him, described the same dynamic as an established practice among competitive rowers. Asked what he would change, he explained why the numbers are sometimes removed deliberately in training: "there are times where we tape over that screen," because "we were trying to get a full aerobic workout without pushing our body into the anaerobic," and "it's really easy to… be like, I'm not

pulling hard enough when you see those numbers." On the console itself he was blunt: "the feedback that you get is valuable, but the way it's presented… hasn't changed very much since the 1980s."

A novice on first exposure and an expert of three decades independently describe a precise numeric display as exerting a pull toward optimizing the number. That convergence is the strongest support in the corpus for the argument in Section 2.4 that unambiguous performance feedback carries evaluative pressure alongside its information.

A second displacement recurred, and is not straightforwardly a success. Participants described the display removing them from the experience of effort: P5, "helped me to divert my mind from the tiredness"; P6, "I didn't feel tired at all," and, comparing the session with habitual training, "when I'm doing the rowing in the gym, I get more tired than I got tired here"; P14, "row without really feeling it that much." P16 described the same effect in terms of duration: "I didn't realize that 10 minutes is up, it is faster than I thought."

Two accounts in this group are more specific than relief, and both point at adherence rather than comfort. P1 located the effect in the substitution itself — rowing "instead of looking at the erg dashboard," and "sitting in boring room like a gym environment," so that "the simulation helps to distract from that fact, which makes it less physically exhausting and gives some mental relief." P11 put it as a counterfactual about his own behaviour: the visuals "made me continue rowing," and in a gym "I'm not sure if I was going to continue rowing this much and I would continue rowing that fast." That is a claim about work done rather than about how the session felt, and it is the kind of claim an adherence study would be built to test. The questionnaire points the same way from the other side: "the projected visuals made rowing feel more tiring than usual" returned the second-lowest mean and the second-lowest variance of any item ($M = 1.12$, $SD = 0.33$), behind only the rhythm-disruption item on both, so no participant experienced the display as adding to the cost of the work.

P4 named a quality the other accounts leave implicit. Asked whether the visuals were ever distracting, he grounded the answer in the coupling rather than in the imagery: the display was "relevant to my rowing action so it's like the effect of what I'm doing," and therefore "it isn't overwhelming or anything. It's like natural when I'm rowing." Legibility here is not a property of the picture but of the correspondence — a display that answers what the rower is doing does not compete for attention, which is the mechanism Theme 1 describes and the condition Section 2.2 argues peripheral uptake requires.

One elite participant reports the opposite, and the counter-case matters more than its single instance suggests because it comes from the group whose training the displacement would most plausibly harm. P17 described the display as pulling attention *toward* the task rather than away from it: "when I'm erging in the wintertime, otherwise I might have music

going or something. It's pretty easy to start thinking about other stuff when you're on the erg, but with this, it kind of brought my attention more to what I was doing on the machine." Where novices described the display as relief from effort, this rower described it as a check on mind-wandering. Both are attentional effects and they point in opposite directions, which suggests the effect may depend on what the rower's attention would otherwise have been doing. Chapter 6 returns to whether that is desirable in a training context.

The most precise articulation of the attentional effect came from P12, and it arrives from inside rowing's own coaching vocabulary rather than from HCI. Describing what the display made possible, P12 recalled that "sometimes coaches ask you to see everything when you're rowing, but focus on nothing," elaborating that "you see it all, but you don't focus on anything," and concluded: "And so I found it very easy to do in this situation." The same participant reported that his habitual erg strategy is to shut the visual channel down entirely — "when I work out on erg, I generally close my eyes" — and that during the session "I had to keep forcing myself to keep my eyes open." An expert whose default response to the ergometer's visual environment is to close his eyes described the projection as something he could see without attending to. Asked separately, in writing, what he liked most, the same participant named the quality the design was aiming at before naming the mechanism: "The enhanced calmness, and the ability to *see everything but focus on nothing*, which is what some coaches direct the rower to ideally do from a visual axis perspective."

Theme 4: Interpretation was easy globally but uncertain locally

The low mental-effort scores are qualified rather than contradicted by the interviews. Specific visual elements went unmapped. P15 was unsure whether recurring shapes were waves: "it took me a second to interpret." P7 was initially confused by droplets before resolving them: "the more I saw the droplets… I was able to figure out, okay, it's my being in, it's just me being imbalanced on my handle." P14 could not stabilize the relationship between pull force and flow speed: "I got kind of confused with how is the pulling motion related to the speed of the rowing or speed of water."

Global legibility was high while specific features remained locally opaque, and participants differed in whether they resolved the opacity through exploration. Notably, nobody reported this as an obstacle: the same participants who could not decode a particular element rated interpretive load at or near the floor.

P7's case is the most interesting, because what he eventually decoded was not designed to be decodable. The droplets that resolve into a signal of handle imbalance let him report

knowing "when you're out of balance," which is technique-relevant information extracted from a display that was never intended to teach technique.

Theme 5: Stopping was possible but not commandable

The flow field decays when rowing stops: momentum dissipates and the surface gradually slows, exactly as a hull loses way. What it does not do is stop *on command*. Bringing the display to rest requires waiting out the decay, and two participants at opposite ends of the expertise range identified that gap independently.

P8 named it as what he liked least: "I would have liked to… stop the boat and then start again just to sort of… get a feel for that," observing that "the water is always moving past at some speed" so "it felt like I wasn't really affecting a boat at all." P10 arrived at the same point through rowing practice and named the real manoeuvre: "Giving the ability to, like, hold water. Right? So, on the water, if we wanted to stop the boat, we would, like, put our blades in the water and hold water, and the hull would stop." On the water, holding water is an *action* that arrests the boat; in RowSim the only route to rest is to stop rowing and wait. He wanted it for a specific purpose — "there was a couple times where I wanted to do starts" — and his workaround was to let the field run down before each attempt, which is visible in his session as the two long pauses reported in Section 5.3.

P5 identified the complementary limit from the other side, wanting motion to persist more strongly when not rowing so that the boat still reads as carrying way: "there was no pattern when I'm not rowing."

Together these describe a display that models the passive loss of momentum but affords no active braking. The vocabulary covers effort and its decay, and not the deliberate arrest of motion — which matters because in rowing that arrest is itself a skilled action. That an expert asked for it by its technical name, and a first-time rower asked for it as a way to feel agency over the boat, suggests the missing element is a commandable rest rather than any particular resting behaviour.

Theme 6: Participants predicted their own novelty decay

Unprompted, five participants (P1, P5, P6, P13, P16) anticipated that engagement would not last. P5 was most explicit: "the same patterns again and again, definitely overwhelmed after some time because I did not want to look at it after 5, 6, 7 minutes. So yes, for a long run, same pattern, it's not suitable for me." The same participant had already said of the motion, "I see the same pattern every time, which makes me a little bored after some time." P1: "if

it's the same thing for a very long time… it could feel repetitive." P6 suggested variation "after 5 or 10 minutes of rowing."

P13's version is the most measured and comes from someone who ergs regularly: "it's only 10 minutes, that's like a warmup. So, if we did a whole hour-long, like a longer workout, maybe there would be a different perspective on this question." P12 independently pointed at the same duration gap from the opposite direction, saying that for "my erg winter workouts, which are kind of long (40 minutes of erging or more), this would enhance that experience."

P16, who wanted a larger projected area, put the mechanism plainly: "right now, we have the projected visual of the area right around the machine, and it's small. And I feel like after a while, it can still be monotonic."

What participants are describing has a name outside HCI. Hedonic adaptation is the well-documented tendency for the affective response to a constant stimulus to diminish with repeated exposure, so that what is initially pleasurable becomes neutral without any change in the stimulus itself (45). Later work qualifies the strong form of the model — adaptation is neither complete nor uniform across people or domains, and its rate varies with the kind of stimulus (35) — which is precisely why the question is empirical for a system like this rather than settled in advance. RowSim's situation puts the question sharply: the palette and the mapping were both held fixed, and four participants had substantially explored the visual vocabulary within ten minutes. What is being adapted to is therefore not only the aesthetic but the informational content. How much remains once the mapping is solved is exactly what a longer exposure would establish, and these sessions cannot.

That participants volunteered this within a single ten-minute exposure is the strongest evidence available here on the durability of the effect, and it is evidence from inside the session rather than an inference about it. It is also the one theme that cuts directly against the design position.

5.7 Integration of the Three Strands

The three strands of evidence in this chapter — questionnaire scores, logged behaviour, and interviews — were analysed separately and are brought together here. Table 5.6 sets out where they converge, where one qualifies another, and where a strand carries a finding the others cannot reach.

The last of those is the point worth stating plainly, because it is easy to read a mixed-methods chapter as though the numbers were the findings and the interviews were illustration. Three of the seven rows below run the other way. On pausing, the behavioural log is not

merely incomplete without the interviews — it is *misleading* without them, because the same metric records fatigue in one group and deliberate structure in the other. On absorption, the interviews account for a quantitative anomaly, the widest item spread in the study, that the questionnaire data alone can only report. On novelty decay, there is no quantitative evidence at all, because a single session cannot produce any; the sharpest challenge to the design position in this thesis comes from five participants who volunteered it unprompted.

Two consequences follow for how the rest of the thesis should be read. First, where the strands converge, they are not independent confirmations of one another: participants who enjoyed the session are likely both to score it highly and to speak well of it, so convergence raises confidence modestly rather than decisively. Second, where they diverge, the divergence is the finding. The pausing row is the clearest case in this study of a behavioural measure that cannot be interpreted without asking the person who produced it.

Table 5.6: How the three strands of evidence relate on each principal finding. The right-hand column states what can be claimed once all three are considered, which in three cases is not what any single strand would support on its own.

Finding	Self-report and logs	Interviews	Integrated reading
Engagement was high	Adapted UES-SF 4.37 (SD 0.43); REQ 4.54 (SD 0.32)	Sustained interest described across both groups	*Convergent.* Self-reported engagement was high and the accounts match the scores.
Engagement did not separate the groups	No engagement measure differs detectably (all $p \geq .195$; $r_{\text{rb}} \leq 0.42$)	Both groups describe the display in similar terms	*Convergent.* No large group difference; smaller differences remain possible at these group sizes.
Rowing separated the groups decisively	All seven physical measures; three at $r_{\text{rb}} = 1.00$	Elite participants discuss training context and metrics	*Convergent*, and independent confirmation that recruitment-time grouping was real.
The display did not interfere	Rhythm item at the floor for sixteen of seventeen	Little reported distraction; one expert reports peripheral use in coaching's own terms	*Convergent on perceived non-disruption.* Neither strand measures attentional uptake (Section 6.1.2).
Pausing	$\rho = -0.57$ with engagement among casual participants ($p = .054$)	Casual accounts describe fatigue; the one elite pauser describes self-imposed pieces	*Divergent, and the interviews are decisive.* The same behavioural metric denotes different states by group; the log alone would misread P10 as disengaged.
Absorption	Focused attention is the only subscale off the ceiling; "lost myself" returns the widest spread in the study	Participants distinguish being absorbed *in* the task from losing themselves	*Qualitative explains quantitative.* The spread is a property of the item, not of the sample.
Novelty decay	No measure available: a single session cannot show adaptation	Five participants predict it unprompted (P1, P5, P6, P13, P16)	*Qualitative only.* The strongest challenge to the design position rests entirely on participant accounts.

5.8 Colour and the Refusal of Realism

Colour was the most-discussed single property of the visualization: twelve of seventeen participants raised it unprompted. Because the palette was a deliberate design decision rather than an incidental setting (Section 4.1), the responses are reported here in their own right. It is worth being clear about what was fixed and by whom: RowSim exposes both the fluid colour and the background as freely selectable, with an optional cycling rate, so any of the palettes participants asked for could have been supplied in seconds. Holding one palette constant across all seventeen sessions was a decision about the study, not a limit of the system. What participants are responding to below is therefore a research choice, and their responses are evidence about that choice rather than about what the software can do.

Eight measured the magenta against a water referent (P3, P4, P5, P8, P9, P13, P15, P17). P4 found the "pinkish colour wasn't relevant compared to water." P8, more practically, found it hard to see: "the pink is just hard to see… it blends into the white a little bit." P13, who has a visual-arts background, framed it as legibility rather than realism: "this I find is so faint that I'm straining to see what it is," wanting "more contrast and a darker color." P17 treated it as an opportunity rather than a fault, wondering "if there are ways to simulate water colors maybe a bit more, whether it's a sunny day, it would be more bluie-ish, or overcast day, maybe grey-ish water" — a request not for accuracy but for the palette to carry a further variable. Three others (P1, P2, P6) raised colour with no reference to water at all, asking only for variety over time.

Three of the twelve who raised colour have first-hand experience of rowing on water: two elite participants (P13, P17) and one casual participant who rows on water (P8). All three raised it, and none of them raised it as a preference — each described a specific perceptual consequence. That is worth setting beside the one engagement measure that came closest to separating the groups: aesthetic appeal, at 4.47 among casual participants against 3.93 among elite ones ($p = .195$). Participants with a referent for what rowing water actually looks like were the more exacting audience for a display that declines to look like it.

One participant articulated the design rationale unprompted, and more clearly than the design documentation had. P11 valued the non-realistic palette precisely because it declined to imitate: "having a different colour than blue was actually pretty nice. Like it felt like, yeah, I was in an environment, but it wasn't necessarily at the sea." The distance from realism was what made the experience legible as its own thing: "that difference made it clear that I was here playing in XR game rather than in reality." P11 returned the joint-highest engagement scores in the study (UES 5.00, REQ 5.00) and never paused.

The distribution of responses is itself informative about what the abstraction did. Participants did not fail to see water; the fluid dynamics were legible enough that a water frame was imported by almost everyone, and the colour then refused it. That is a productive tension rather than a failed illusion — the display was realistic enough in behaviour to summon a referent and unrealistic enough in appearance to decline it, which is the condition Section 2.3 describes.

P13 drew the natural conclusion, proposing as future work exactly the experiment this study deliberately did not run: "I think it'd be really interesting to think about how the colour affects the experience... it would just be interesting to see what the responses are based on the colour even."

Beyond colour, five participants raised sound (P2, P3, P4, P7, P11) and eight wanted projection beyond the floor (P5, P6, P7, P11, P12, P13, P16, P17). P17 tied the request to the wake idea directly, wanting "a screen in the back to present where you've been as you leave, as you row." P2 was most direct on sound: "The lack of sound, of the water, was a bit immersion breaking." P11's case runs the other way and is more revealing: the ergometer's own flywheel noise was mistaken for the visualization's, "I thought it was a water droplet sound... but yeah, I was fooled" — attributing the impression, in the same breath, to the environment the projection had created. An unmodified mechanical sound was absorbed into the fiction because the visuals made it plausible.

5.9 Comfort and Accessibility

Two participants disclosed relevant pre-existing conditions, and their outcomes ran in opposite directions.

P8 had disclosed susceptibility to motion sickness on the background questionnaire and reported mild symptoms during the session: "I had minor amount of motion sickness if I looked at them for too long, but I don't think that is a concern overall. I can just look away, and it resolves itself." P8 also took seventeen pauses, more than any other participant. The self-report and the log agree, and both point the same way.

P12 had disclosed chronic vestibular dysfunction and reported no difficulty at all: "I have vestibular problems with balance and equilibrium, whatever. It didn't challenge that in any way."

These are two individuals and support no general claim about vestibular safety. They are reported because they are the only direct evidence the study contains on a question that matters for a floor-projected moving display, and because the pair is more informative than either case alone: susceptibility to motion sickness and vestibular dysfunction are not

the same condition, and this display appears to interact with them differently. That P8's discomfort resolved on looking away also bears on the peripherality claim, since a display that could be disengaged by a glance elsewhere was not compelling attention.

Chapter 6

Discussion

6.1 Interpretation of Findings

6.1.1 Engagement did not track a large difference in rowing

The central result is a juxtaposition rather than a single number. Elite and casual rowers differed by a factor of 3.7 in mean power, 4.3 in work per stroke and 2.6 in peak drive force, with every one of seven physical measures separating the groups and five doing so at $p \leq .0006$. On overall engagement they differed by 0.21 scale points, on the REQ composite by 0.09, and on no engagement measure significantly.

The immediate objection to a null is that five elite participants cannot detect anything, and the design answers it internally. The same test, applied to the same seventeen people, separated the groups decisively elsewhere in the same table. Whatever the comparison lacked, it did not lack sensitivity to large group differences: it found them wherever they existed. The comparison is demonstrably capable of detecting large between-group differences in this sample, so the absence of a detectable engagement difference is informative rather than vacuous. It bounds the effect rather than abolishing it: a difference of the size that separates these groups physically is not there, while a smaller one could be and would not have shown at these group sizes.

What that result says is bounded and worth stating carefully. It does not say that expertise is irrelevant to how ambient feedback is received — the item-level differences in Section 5.4 show that it is not. It says that the *overall level* of engagement did not depend on how well or how hard the participant rowed.

This matters because it is not what a performance-linked display would predict. A system whose appeal derives from displaying achievement should be more rewarding to those achieving more; RowSim's was not. The interviews suggest why. What participants valued was the coupling itself — the fact that something in the room responded to them — and coupling is available at 8 W as readily as at 166 W. Responsiveness is not a scarce resource distributed in proportion to power output.

Two qualifications follow it. Engagement scores clustered near the top of both instruments, and a compressed range is easier to find no differences in; the item-level analysis mitigates this but does not remove it. And the elite group is five people, so the finding is that no difference appeared between these five trained rowers and these twelve casual ones, which is a smaller claim than a population statement. The nearest thing to a difference is aesthetic appeal ($p = .195$), where the trained group rated the display lower — consistent with Section 5.8, in which participants with first-hand experience of water were the more exacting audience for a display that declines to look like it.

6.1.2 The display was reported as non-disruptive, and an expert named the mechanism

The clearest quantitative finding is a near-floor score on rhythm disruption (Section 5.2), and it is easy to overlook because floor scores rarely look like results. This one is. A moving display placed in the visual field during physical exertion could plausibly have drawn focal gaze or disturbed the catch; no participant reported that it did, and several described using it to hold a rhythm rather than fighting one.

It is worth separating three things that this section otherwise risks running together. What was *measured* is perceived non-disruption: participants reported, with the least variance of any item in the study, that the display did not interfere with their rowing. What is *inferred* from that, together with the interviews, is compatibility with peripheral use. What was *not measured* is attentional uptake — whether peripherally available information was actually picked up. Low reported disruption is consistent with successful peripheral presentation, and equally consistent with participants looking directly at the projection and not minding.

Precision matters here: what would actually settle this, because the obvious answer is not the right one. Eye tracking records foveation, and peripheral perception is by definition what is not foveated; a gaze record showing few fixations on the floor would be compatible with the display being used peripherally and equally compatible with its being ignored, and the two are exactly what needs separating. What distinguishes them is whether peripherally available information was *picked up*, which is a question about perception rather than about pointing direction — and that calls for a dual-task or peripheral-detection paradigm, in which a change is introduced in the projection and detection is measured while the primary rowing task continues. Section 6.4 takes this up.

Between those two positions sits the most striking single piece of evidence in the corpus. P12, an elite rower, described what the display made possible using rowing coaching's own formulation for peripheral attention: coaches "ask you to see everything when you're

rowing, but focus on nothing," such that "you see it all, but you don't focus on anything." His verdict was that "I found it very easy to do in this situation."

This is not a participant agreeing with the thesis; it is a participant arriving at the same construct from a different tradition and applying it unprompted. Weiser and Brown define the periphery as "what we are attuned to without attending to explicitly" (128); competitive rowing, with no knowledge of calm computing, instructs its athletes to see everything and focus on nothing. That two vocabularies developed for entirely different purposes converge on the same distinction is qualitative evidence that the distinction is meaningful to at least one experienced rower, and P12's report is consistent with the display remaining available to him without becoming a competing focal task. It is one participant's account of his own attention, not a measurement of where his attention went.

The same participant supplied the sharpest statement of the problem RowSim addresses. Asked what stood out, he mentioned that his habitual practice on the ergometer is to close his eyes, and that during this session "I had to keep forcing myself to keep my eyes open." An expert rower's default response to the ergometer's visual environment is to shut it out. The display gave him a reason not to.

None of this is gaze measurement, and it remains self-report. But it is self-report from a participant with three decades of relevant expertise, using a technical vocabulary he brought with him rather than one the study offered.

6.1.3 Ambiguity invited exploration, and the mechanism was mostly calibration

A majority of participants probed the mapping rather than seeking a legend, and interpretive load stayed at the floor of both instruments (Section 5.6). The ambiguity-as-resource position of Section 2.3 is therefore supported on its own terms: an under-specified display invited engagement rather than producing frustration.

The mechanism, though, is mostly not the one Gaver et al. describe. Their claim is that ambiguity prompts users to construct personal meaning (47). What most participants did was converge on a shared, broadly correct account of what drove what — P6 systematically comparing three techniques and concluding that elbows-down was fastest, P10 stress-testing the mapping's limits at different rates. That is calibration, and calibration terminates. It explains both why participants found the display absorbing and why five of them predicted their own boredom within the session: a puzzle that can be solved is a finite resource. It also has a name outside HCI — hedonic adaptation, the diminishing affective response to an unchanging stimulus (35, 45) — and a mapping that can be exhausted removes the

informational novelty alongside the aesthetic one, which is a sharper version of the usual concern rather than a separate problem.

P15 is the exception that clarifies the rule. Her account of the colour shifting from pink to white as the handle moved, reasoned through an analogy to how water changes colour with viewing angle, is a personal interpretive model of a mechanism the system does not implement. That is meaning-making in Gaver's sense, and it was rare.

What was calibrated, though, was not the whole mapping but its loud part. Five rowing quantities drove the display continuously (Section 3.10), and the interviews are almost entirely about two: the effort that widened and brightened the wake, and the timing that placed it. Not one participant mentioned the bloom that tracked stroke consistency, or the slow warming that followed sustained rhythm, though both ran in every session. The mapping that was solved, in other words, was the part of it loud enough to be noticed being solved.

One exchange makes the threshold concrete. P2 proposed, as an improvement, that the visuals should change colour with speed — "if I'm going faster, maybe the visuals change for a whole different colour… I think it would be a good visual feedback." RowSim already does that; the mode was switched off for the study (Section 4.1). A rower asked for a coupling that was sitting in the system, unbuilt only in the sense of being disabled, which is a cleaner demonstration than any instrument that the couplings participants can name are the couplings the design made loud.

That cuts two ways for the novelty argument. The exhaustible resource is smaller than the full coupling set, which shortens its life: a rower who has solved the effort and timing mappings has solved what they can see, and the remaining structure will not hold them. But it also means the ceiling observed here is a ceiling on *this* configuration rather than on the approach. Two couplings ran for ten minutes without anyone noticing them, which is evidence that a display of this kind has range left — and Section 6.4 sets out the experiment that would use it, which is to raise those gains until participants can report the effect and see whether the evaluative pressure Section 2.4 warns about arrives with it.

There is a methodological consequence in this too. Participants' accounts of what drove what are demonstrably partial: they cover the couplings above threshold and are silent about the rest, so an account's completeness cannot be assumed from its confidence. Convergence is the check. Several participants independently reaching the same reading is hard to produce by accident, whereas one vivid observation is not, and that argues for weighting agreement across participants above the striking individual report — the opposite of the instinct a good quotation encourages.

The distinction matters for design. If ambiguity engages through calibration, it is spent; if it engages through meaning-making, it renews. RowSim mostly did the former. Gaver's own examples — ambiguous artefacts in domestic and cultural settings — differ from a real-time action–effect system in exactly the respect that matters: a display that responds to you within 100 ms invites you to work out how, whereas one that does not respond invites you to work out what it means.

6.1.4 Displacing effort is not the same as supporting training

Participants repeatedly described the display removing them from the experience of effort: "helped me to divert my mind from the tiredness" (P5), "I didn't feel tired at all" (P6), "row without really feeling it that much" (P14). P4 reported a ten-minute session feeling like two or three.

Two things need separating here, because only one is settled by the literature this thesis draws on. Wulf's work concerns where attention is *directed*: an external focus, on the effect a movement produces in the environment, generally benefits motor performance and learning relative to an internal focus on the body (133, 134). On that criterion RowSim is well designed. A floor projection that renders the consequence of a stroke as displaced fluid is an external-focus cue almost by construction, and the participants who calibrated force against the visible result were doing what that literature recommends.

What participants also described is different, and the external-focus literature does not speak to it: not attention redirected from the body to its effects, but attention withdrawn from exertion altogether. Whether that is benign or costly for a training tool is a question about associative versus dissociative attending in endurance activity, which this study did not measure. The negative association between focused attention and work per stroke ($\rho = -0.43$) points in the concerning direction without settling it.

The elite data sharpen the question, and one elite participant answers it in the opposite direction. P17 reported that the display pulled his attention *toward* the machine rather than away from it: on the ergometer he would otherwise "have music going or something" and finds "it's pretty easy to start thinking about other stuff," whereas with the projection "it kind of brought my attention more to what I was doing on the machine." That is the reverse of the novices' account, and it is the group in which the displacement would matter most.

The two readings are reconcilable, and the reconciliation is the more useful finding. What the display displaces may be whatever the rower's attention would otherwise have occupied. For a novice on an ergometer, that is the discomfort of unfamiliar exertion, so the display reads as relief. For a trained rower doing a familiar winter session, it is whatever

fills a long and unremarkable piece of training, so the same display reads as re-engagement. The intervention is constant; the baseline is not. This is testable, and the demographic questionnaire already carries the variable it needs: the prediction is that the direction of the attentional effect tracks prior ergometer habit rather than competitive expertise. What this sample cannot do is separate them. Every elite participant reports habitual ergometer use and only one casual participant does, so the two candidate explanations coincide across sixteen of seventeen cases. Recruiting casual rowers who erg regularly, and competitive rowers who train mainly on water, would pull them apart; nothing short of that will.

All five elite rowers gave the minimum response to "the experience required a lot of mental effort" and the highest group mean to "Overall, I enjoyed the experience," yet were the more reserved group about substituting the display for their console. P10 explained why in operational terms: competitive rowers sometimes "tape over that screen" during aerobic work, because "it's really easy to… be like, I'm not pulling hard enough when you see those numbers." That is an athlete describing a deliberate, practised removal of numeric feedback in order to protect a training intention — which is the same manoeuvre RowSim performs by design, arrived at independently by practitioners.

It also cuts both ways. Rowers tape over the display when they want less evaluative pressure, and read it when they want to hit a target. A display that cannot be read when precision is wanted has traded one problem for another, which is precisely why P10 asked for the PM5 data back with "the ability to show or hide whatever you want to see." The honest reading is that RowSim demonstrably shifts attention toward movement effects, which is well supported as beneficial; whether it additionally suppresses interoceptive signal an athlete should attend to remains open, and a workload or attentional-focus instrument rather than an engagement questionnaire would be needed to close it.

6.1.5 An expert convergence on the problem with numbers

One convergence deserves separate notice because it arrives from both ends of the expertise range and neither participant was asked about it.

P9, a casual rower on first exposure, observed that "when I was usually looking at a metric, I'm sort of optimizing for the metric in particular. But when I do not have a metric, I'm just sort of having fun with it to some extent." P10, after more than thirty years of competitive rowing, described the same dynamic as an established coping practice and gave the reason his crews tape over the console.

Chapter 2 argued on theoretical grounds — from self-determination theory — that unambiguous performance feedback supplies competence information and evaluative pressure at

the same time, and that the two can pull against each other (101, 115). Two participants who share no relevant background independently described experiencing exactly that trade-off. This is the strongest empirical support the study offers for a claim the thesis had to make theoretically, and it was volunteered rather than elicited.

The design has a partial answer to this already, though a quieter one than the theory would want. Consistency is itself competence information, and RowSim carries it: stroke regularity drives bloom intensity, and sustained rhythm warms the field, so a rower holding a steady rate is shown something without being told a number. Both effects are subtle by construction — a gradual brightening and a slow shift in warmth, at amplitudes that sit inside the visual noise of a moving fluid — and no participant named either. That subtlety is the design's answer to the trade-off: a competence channel quiet enough to carry no evaluative pressure. Whether it is quiet enough to carry no competence information either is the question, and it is answerable by experiment rather than by argument — raise the gain until participants can report the effect, and see whether the pressure arrives with it.

6.1.6 Stopping was not something the rower could do

Chapter 2 opened on a perceptual claim rather than an interface one. On the ecological account, self-motion is not inferred from cues but specified directly by the lawful transformation of the optic array at a moving point of observation (48), and the tradition following it has shown that people use that transformation to steer and to hold themselves upright, continuously and without deliberation (70, 126). The ergometer's anomaly, on that account, is not that it lacks a display. It is that it supplies every proprioceptive and metabolic sign of vigorous locomotion while the optic array stays still.

RowSim puts motion back into that array. What two participants noticed is that putting motion back is not the same as restoring the coupling.

P8 and P10 arrived at the same observation from opposite ends of the expertise range and for different reasons. P8, a casual rower, said what he liked least was that he could not "stop the boat and then start again just to sort of, like, get a feel for that," and gave the reason plainly: "the water is always moving past at some speed," so "it felt like I wasn't really affecting a boat at all." P10, an elite rower, described the same gap in the vocabulary of the sport — on the water, "if we wanted to stop the boat, we would, like, put our blades in the water and hold water, and the hull would stop."

The ecological frame predicts exactly this complaint, and predicts that it would be about *agency* rather than about realism. What specifies self-motion is not motion in the array; it is motion that covaries lawfully with what the observer does. Flow that continues after

you stop specifies motion you are not producing. A rower who cannot arrest the flow is, perceptually, a passenger.

The system's behaviour matches the complaint closely, and the mechanism is specific enough to state exactly. Two things on screen decay, and they decay on very different timescales.

The *wake* — the ink and velocity a stroke injects locally, and the channel that actually carries effort — is multiplied by 0.97 every frame. At sixty frames per second that is a half-life of 0.38 s: a stroke's visible contribution is more than four-fifths gone one second after it lands. The *carrier flow* — the field-wide drift that specifies self-motion — is governed by nothing but quadratic drag between strokes, $\dot{v} = -C_d\, v\,|v|$ with $C_d = 0.04$, whose solution is

$$v(t) = \frac{v_0}{1 + C_d\,|v_0|\,t}\,, \tag{6.1}$$

a reciprocal rather than an exponential. Its half-life is $1/(C_d|v_0|)$, and the shape of the curve matters more than the number. A reciprocal falls fast at first and then flattens: most of the carrier speed is gone within a few seconds, and what remains drains away slowly, so the field does come to rest if the rower simply waits, but the last visible drift outlasts the stroke that produced it by a wide margin. The controller snaps the carrier to zero once it falls below a small idle threshold, and the fluid solver's own dissipation takes the remaining motion out of the field, so rest is reached rather than merely approached.

The asymmetry is what P8 was describing, and it is sharpest exactly where he met it — mid-session, while still rowing. Each stroke's local contribution is gone within a second, but each stroke also tops the carrier back up, so the drift he was complaining about never got the uninterrupted stretch it would have needed to decay. The water is always moving past at some speed" is an accurate report of a rower's experience of this system, even though the same system settles to rest if left alone. The channel that answered him disappeared almost at once; the channel telling him he was moving faded on a timescale his rowing kept resetting. That inversion is the whole of the complaint. And the rower has no remedy inside a session, because the impulse the controller applies is clamped to increase speed only: there is no negative-force term anywhere in the physics, and therefore no input — no handle position, no hold, no reversal — that subtracts carrier velocity. Stopping, in RowSim, is not an action; it is the absence of action, resolved by decay. On the water it is an action: holding water is something a rower *does*, with an immediate and proportionate consequence. The design supplied the first half of the coupling — flow responds to effort — and omitted the half that makes the flow legible as one's own.

This reframes what the thesis found. The headline result is that a peripheral, non-prescriptive display was habitable during sustained physical work, which is a claim about attention. The carrier-flow observation is a claim about something else: that restoring optic flow to a stationary exerciser is a coupling problem, not a rendering problem, and that the coupling is incomplete unless deceleration is as actionable as acceleration. Nothing in the ambient-display literature predicts this, because that literature concerns observers who are not themselves the source of the motion being displayed. It follows instead from taking the ecological framing seriously as a design constraint rather than as motivation.

It also identifies the cheapest available improvement. The obstacle field discussed in Section 6.4 is one route — something to act *on* rather than merely *in* — but the more direct one is a braking input: a sustained handle position, or a detected hold, that arrests the flow proportionally rather than waiting out the decay. Equation 6.1 names the change precisely — remove the clamp that keeps the applied impulse one-signed, so that a hold can subtract carrier velocity the way a stroke adds it. That is a few lines in the physics controller, and it would convert the display from something that responds to effort into something that answers to action. Whether it changes the reported experience is testable, and the two participants who identified the gap are the ones whose accounts would decide it.

6.1.7 The open task made behaviour interpretable

Because participants were given no target and were free to stop, the logs record choices. Thirteen of seventeen never paused, and the four who did divide cleanly into two kinds once the interviews are consulted: three casual participants whose pausing accompanied the corpus's clearest reservations, and one elite participant whose two rests bracketed self-imposed training pieces.

Two features of this deserve emphasis. First, the correspondence between account and behaviour is predictive rather than merely consistent: P7 and P11 volunteered that the display kept them from taking breaks, in sessions where neither could see the log, and neither has a recorded pause. Self-report and behaviour agree on a measure neither participant knew was being taken. Second, the same behaviour carries opposite meanings in the two groups, and only the qualitative strand distinguishes them. A quantitative analysis alone would have averaged an expert's interval training together with a novice's fatigue and produced a number describing neither.

This is the methodological argument for the open task made concrete. Had the protocol specified ten minutes of continuous rowing at a target rate, every participant would have complied approximately, the pause measure would not exist, and P10's self-direction — the

single clearest expression of expertise in the behavioural data — would have been prohibited by the instruction.

6.1.8 Two engagement modes, and an instrument that cannot see them

The divergence between the two focused-attention items (Section 5.2.2) is a measurement observation with a substantive reading behind it. Participants distinguished being absorbed in the activity from losing themselves, and four scored these items three or more points apart.

The interviews suggest two ways of engaging. Some participants engaged *immersively*: P4, P11, P14 and P15 describe flow-like absorption, time distortion and being transported elsewhere. Others engaged *analytically*: P6, P8 and P10 describe curiosity-driven experimentation while remaining fully self-aware, with P8 explicit that he was "purely trying to figure out how what I did affected the visuals." Both groups rated the system highly; neither was disengaged. But only the first is what an absorption subscale is built to detect.

The implication for evaluation is direct: a single absorption construct will systematically under-credit the analytic mode, and an instrument built on it will report lower engagement for exactly the participants who interacted with the system most deliberately.

6.2 Comparison to Literature

6.2.1 Does calm computing survive exertion?

Section 2.2 closed on a caveat Weiser and Brown make against their own programme: "Not all technology need be calm. A calm videogame would get little use; the point is to be excited" (128). Rowing at training intensity has more in common with that exception than with the quiet offices in which calm computing was worked out, and whether principles derived from the periphery of a settled attention transfer to the periphery of a body under load was left there as an open question.

On the disruption criterion, the transfer holds. Weiser and Brown's premise is that peripheral information can be "informing without overburdening," and Buxton's is that a low-bandwidth persistent channel costs the foreground less than a high-bandwidth bursty one (25). A continuous floor projection during exertion did not register as a burden: the rhythm-disruption item returned the lowest variance in the instrument, and participants preferred the projection to a standard rowing display ($M = 4.59$). Physical load did not, in this sample, convert a peripheral display into a competing demand. That is the finding this

thesis can most defensibly add to the ambient-display literature, because it is the condition that literature has least often tested.

The test was also a demanding version of the question, because the console was covered for the recorded segment (Section 4.3). Participants were not choosing the projection over an available numeric read-out; they had the peripheral channel and nothing else, for ten minutes of physical work, with no visible clock. That the resulting sessions were largely uninterrupted, and that nobody reported the display as insufficient to row by, is a stronger result than it would have been with the console left on — a peripheral display was not merely tolerable alongside a precise one, it was habitable in its absence.

On the criterion Weiser and Brown themselves treated as definitive, this study has less to say, and the reason is the same design decision. Their claim was never that calm technology remains in the periphery; it was that it should "move easily from the periphery of our attention, to the center, and back," and that "the individual, not the environment, must be in charge" of that movement. With only one channel present, the traffic between two of them could not be observed. What the interviews do show is movement within the single channel: the probing behaviour in Theme 2 is recentring under participant control, P8's motion sickness resolving on looking away is the reverse move, and P12's account of seeing everything while focusing on nothing describes the resting state between them. P10 then asked for precisely the missing arrangement — the PM5 data restored with "the ability to show or hide whatever you want to see" — which is Buxton's transition condition requested by a user who had spent ten minutes on the peripheral side of it.

One qualification stops this from being a stronger claim. The probing that supplies the recentring evidence was, on the reading advanced above, a finite calibration: a display recentred only until it has been solved satisfies the traffic condition briefly and may then stop satisfying it.

6.2.2 Ambient displays

Ambient-display research supplies the criteria against which RowSim's peripherality can be judged: a display should remain informative without demanding focal inspection, and should degrade gracefully when attention is elsewhere (9, 90, 94). On the attentional criterion this study is a clear positive case, and it extends that literature into a setting it has rarely addressed. Ambient-display evaluations are overwhelmingly conducted in calm environments — offices, homes, public architecture (44, 117, 130) — where the competing demand is cognitive. Here the competing demand was physical exertion, and peripherality survived it.

On ambiguity the relationship is more complicated, for reasons given above: RowSim's evidence supports the claim that ambiguity productively engages, while suggesting that the mechanism in a real-time action–effect system differs from the interpretive one Gaver describes for slower, more contemplative artefacts.

6.2.3 Mixed reality and exergame research

The engagement levels observed sit alongside a body of exergame work in which embodied, physically active play is used to make training motivating rather than merely instructive: exergames developed for rehabilitation settings are motivated in exactly those terms (51), and interactive floor projection has been used to support physically active experiences in shared spaces (120). It fits too with the framing of exertion games as a design space in which bodily engagement rather than instruction carries the experience (83). The REQ item on reducing ergometer monotony ($M = 4.35$) and the near-unanimous willingness to recommend ($M = 4.82$) sit comfortably in that tradition.

Where this study adds something is on the mechanism of that enjoyment. Much exergame design achieves engagement through explicit gamification — goals, scores, competition (73, 105, 125). RowSim supplied none of these: no score, no target, no opponent, no numeric readout, and a task instruction that explicitly declined to set one. That comparable engagement arose without them suggests that responsiveness alone — a legible, immediate action–effect coupling — may be sufficient, and that goal structures are not the only available route. Where participants wanted goals they supplied them: P11 asked for a destination and checkpoints, and P10 constructed his own interval session. Both are requests for structure, and both name something the current design leaves to the rower.

6.2.4 Rowing and sport technology

The stroke-timing values recorded here place both groups where the biomechanical literature would predict. Casual participants rowed at 23.7 spm with a drive-to-cycle ratio of 0.32 and covered 5.98 m per stroke; elite participants rowed *slower* at 20.9 spm with a ratio of 0.27 and covered 10.82 m per stroke. Length and force per stroke rather than stroke frequency is the established signature of trained rowing (67, 110, 111), and the fact that recruitment-time group assignment is reproduced this cleanly by the ergometer's own telemetry is a useful validation of the grouping.

The more useful comparison is with ARrow (62), the closest system in the literature and the design foil identified in Section 3.2. ARrow delivers explicit stroke-phase and technique cues on a tablet screen, with the coach as its primary reader; RowSim delivers no explicit cue

at all, into the rower's own visual periphery. The difference is not only modality but what each display takes itself to be doing. In Suchman's vocabulary (Section 2.3), a technique cue is a *plan* — a prior specification of correct conduct, against which the rower's action is read as compliance or deviation — whereas RowSim renders the situation and leaves the reading to the rower (118). Suchman's argument is not that plans are useless; coaching cues are among the most useful resources a novice rower can have. The point is that the two systems make opposite bets about where interpretation should sit.

These results establish that the non-prescriptive end of that axis is viable: participants engaged, did not lose rhythm, reported preferring it to a standard console, and in one case extracted technique-relevant information from it unbidden — P7's discovery that the droplet pattern indicated handle imbalance. They do not establish that it teaches anything, because nothing here measured technique acquisition. On current evidence the two systems address different problems.

6.3 Limitations

Single-session exposure. Participants experienced RowSim once. Engagement measured immediately after first exposure to a novel system is difficult to separate from a general novelty effect, and the data make this concrete rather than hypothetical: four participants volunteered within the session that the visualization would become repetitive over longer exposure (Theme 6). The engagement scores should be read as first-exposure values with a plausible downward trajectory.

Group sizes. The comparison rests on twelve casual and five elite rowers. The physical differences are large enough to be detected decisively at these sizes, and the argument in Section 5.4 turns on that asymmetry rather than on the group sizes being generous. But a null on engagement between five trained rowers and twelve casual ones is a statement about these participants; at these sizes only large effects are detectable (30, 69), so a smaller genuine difference in engagement could be present and unobserved. Small-sample constraints of this kind are common to in-person HCI studies: across a full year of CHI papers Caine found a median sample size of eighteen and a modal sample size of twelve (26), which places this study at the discipline's centre rather than below it — but the discipline's centre is still too small to detect a moderate effect.

Ceiling effects. Engagement scores clustered near the top of both instruments, with no participant below 3.42 overall. Compressed variance attenuates correlations with objective measures and is an alternative explanation for the weak associations in Section 5.5 that these data cannot rule out. Item-level reporting mitigates this; it does not remove it.

No direct measure of attentional pickup. The peripherality claim rests on self-report and on design intent, not on an observation that peripherally available information was taken up. As argued above, gaze recording would not settle this on its own, because peripheral perception is precisely what is not foveated; a peripheral-detection paradigm would.

Latency is measured on the input side only. The delay between a stroke happening and the projection changing has two halves. The first — getting the sensor reading into the software — was measured directly from the session logs and is negligible: half of all readings arrived within 1.13 ms of being taken, and all but one in a hundred within 3.64 ms, against the 16.7 ms the system has to render each frame (Section 3.11). The measured input-side delay is well under a single frame interval and is therefore unlikely to be the dominant term in the end-to-end delay.

The second half — rendering the frame and getting it onto the floor — was not measured, because no software timestamp can capture the projector's own delay. What can be said about it is observational rather than instrumented. Nothing in the sessions looked delayed: participants located engagement in the immediacy of the coupling, and the closest thing to a contrary report is P10's "every once in a while, a little bit of a delay of seeing the puddle," which is occasional rather than systematic and which he raised alongside a separate complaint about where the puddles appeared. Every session was filmed with the handle and the projection both in frame, so the measurement is recoverable from material already collected: a frame-by-frame comparison of catch and response would give the end-to-end figure without running anyone again. Until that is done the render half is unquantified, and it is named here for completeness rather than because anything in the results points at it.

Contrast was a constraint, and it is confounded with the design. Two participants reported difficulty seeing the display, and both were specific about it. P8 said "the pink is just hard to see… it blends into the white a little bit." P13, who has a visual-arts background, put it as a legibility problem rather than an aesthetic one: "this I find is so faint that I'm straining to see what it is," and asked for "more contrast and a darker color." Section 5.8 reads these as responses to the palette, which is how they were offered. They are equally consistent with a different cause.

The installation threw a large image from two WXGA units at 1280×800 each onto a floor. Room lights were off throughout, so ambient light was not the problem; the surface was. The lab floor is matte and near-black, excellent at absorbing stray light from the overhead rigs and correspondingly poor at returning a projected one, and four white sheets taped together were laid over it to recover most of what it was costing (Section 3.3, Figure 3.1). Sheeting raises reflectance; it does nothing for a projector's black level, and the seams and creases visible in Figure 3.3 scatter a little of what it gains. The darkest parts of

the image therefore stayed lifted and the whole frame carried less contrast than the source renders in Figure 3.13 suggest. Moderate-opacity magenta on white, at that brightness and pixel density, would be faint whatever the palette. **Nothing in this study separates "the colour was hard to see" from "the projector could not render it,"** and the two prescribe opposite fixes — change the palette, or change the hardware. A single session on a brighter, higher-resolution installation with the same magenta would settle it, and should precede any redesign of the colour.

The same budget bears on features subtler than the fluid. The caustic layer, which was running in every session, is a moderate-amplitude lighting effect and is exactly what a poor contrast budget loses first (Section A.8.1).

Deployment dependence. The experience depends on projector placement, sensing reliability, room lighting and calibration even when the software is unchanged — a general concern for projection-based floor displays (135). P10's observation that the ergometer model's seat height affected the sense of being in a boat, and P12's initial doubt about the projection surface, are both instances of this. A less obvious instance is internal to the software: the simulation grid is allocated once at launch and thereafter stretched to whatever surface is projected onto (Appendix A), so the fluid's spatial detail is fixed by the window at startup rather than by the projector. A larger or higher-resolution installation would not, by itself, produce a finer simulation.

Coupled without being legible. Section 3.10 sets out five rowing quantities that drove the display and a sixth, stroke regularity, that drove bloom without a flag governing it. This complicates the description of RowSim as non-symbolic in one specific way. Nothing was legible as a value, which is the claim the study tests; but harder pulls did make a brighter and wider wake, so a participant could in principle have been calibrating against intensity rather than reading the display in the open way the interviews suggest. Nothing in the data distinguishes these, because no condition varied the couplings. A version with the effort couplings removed — flow driven by stroke timing alone, with magnitude discarded — would separate them, and would be a cheap manipulation to run.

Instrument composition. The REQ is a study-specific instrument without independent validation (Appendix B.2), used descriptively and never as a validated scale. Nine of the twelve UES-SF items were administered with wording set to the rowing context as the scale permits; three were replaced — two by items on perceived control and confidence, one by an item on perceived value — giving a 2/4/3/3 distribution across the four dimensions (Appendix B.1). Item-level means are reported alongside subscale means for this reason.

6.4 Future Work

Measuring peripheral pickup. The central claim about ambient presentation deserves an observation rather than an argument from design intent, and the instrument follows from the claim. What needs testing is whether information was taken up without being looked at, so the measurement should be of pickup rather than of fixation: introduce an occasional change in the projection and record whether it is noticed while rowing continues. That distinguishes a display being used peripherally from one being ignored, which is the distinction the thesis rests on and the one no self-report can settle. P12's account of seeing everything while focusing on nothing is a hypothesis in exactly that form, and it is testable that way.

A commandable rest. Theme 5 identifies a specific, buildable gap. The field already decays when rowing stops; what it lacks is an action that arrests it. Implementing the "hold water" manoeuvre P10 named — a deliberate input that brakes the flow rather than letting it run down — would close the gap and would enable the standing starts he wanted to attempt. Section 6.1.6 sets out why this is a perceptual requirement rather than a convenience: deceleration that cannot be commanded leaves the flow specifying motion the rower is not producing.

A channel the rower does not control. Every quantity currently driving the display is an output of the stroke: watts, force, rate, speed. A physiological signal is different in kind, because it lags effort, is not directly commandable, and reports the body's response rather than its action. Heart rate is the obvious candidate, and the plumbing is nearly free — a paired watch already publishes it, and the ingest path is a HealthKit authorization and one more signal on the transport already carrying eleven (Figure 6.1). The design question is the interesting part, and it is not settled by the ability to read the value. Mapped to intensity it would behave like the couplings in Section 3.10, but with a delay the rower cannot cancel, which is a different perceptual object: a display that answers a minute after you asked. Mapped to something slower — the rate at which the field returns to rest, say — it might read as the water's own state rather than as a readout of the body, which is closer to what this thesis argues an ambient channel should do. It is also the point at which the ethics change, since resting heart rate is health information and the consent language described in Chapter 4 was not written for it.

Duration. A longitudinal study here would not be a robustness check but a test of a specific mechanism. The account in Section 5.6 predicts something falsifiable: if engagement rests on calibrating an exhaustible mapping, it should decay on a timescale set by how long the mapping takes to solve, and a system whose mapping varies should decay more slowly than one whose mapping does not. That is a comparison a multi-session design can run

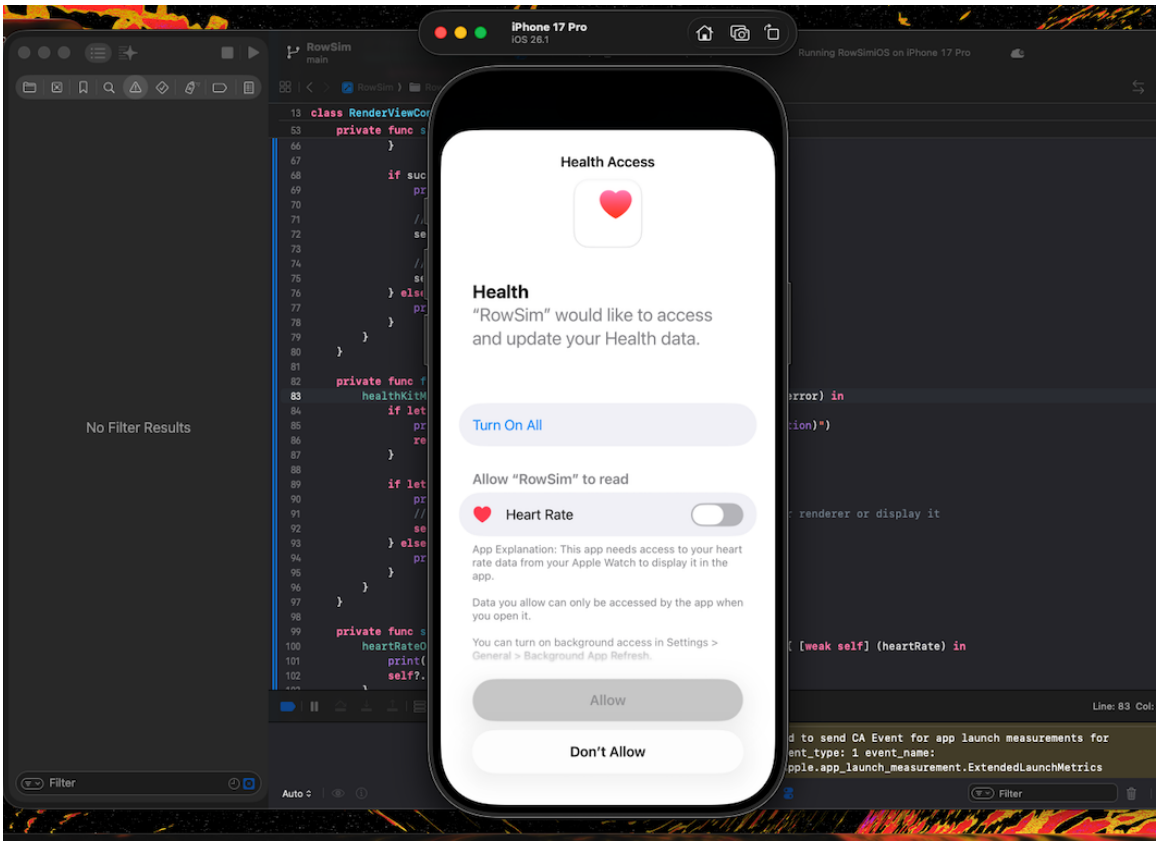


Figure 6.1: Exploratory heart-rate ingest, stopped at the HealthKit authorization prompt. This was a probe rather than a feature: it is not in the published repository, whose iOS target is touch-only (Section 3.4), and the signal is not used anywhere in this thesis or in any study session. The ergometer's own console does report heart rate over Bluetooth, and RowSim logs it, but it is not mapped to the display. The figure is included because it shows the easy half — the transport and the permission flow — while the mapping question discussed above is the hard one.

directly, and P12 named the setting to run it in. He described his winter training as erg sessions of forty minutes or more and said RowSim would make them "go faster" and "be much more pleasant." Those sessions are the ones an ambient alternative exists to improve, and they are four times longer than anything this study observed — so they are also where a mapping that can be solved in ten minutes would first run out.

Colour as an experimental factor. This study deliberately held the palette fixed so it could not become the operative variable (Section 4.1). Having done so, the obvious follow-up is to vary it deliberately, and a participant proposed exactly this: P13 suggested it "would just be interesting to see what the responses are based on the colour even." The system already supports the necessary configurations, including the cycling and speed-mapped modes disabled here.

The visualization as a slot rather than a fixture. The strongest version of that follow-up is not a second palette but a change of unit. Everything this thesis reports is one visualization, held constant across seventeen sessions so that it could be studied. Nothing in the architecture requires that. The fluid solver produces a velocity field, an ink field and a set of stroke events; what turns those into an image is the final visualization pass, and it is separable from everything upstream. A different shader in that slot would be a different display driven by the same rowing.

That reframes the contribution. What was built is closer to a design space than to a product: a sensing and physics substrate with a well-defined interface, into which visual treatments can be plugged. The evidence that this matters is already in the data. Participants did not converge on one preference — P13 wanted more contrast and a darker palette, P17 wanted water colours varying with the weather, P11 valued the magenta precisely because it was not water, and three others asked only for variety over time. A single fixed treatment cannot satisfy that spread, and there is no reason it should have to.

The design question this opens is one of granularity. At one end, a rower chooses a visualization the way they choose a playlist, and the system asks nothing further of them. At the other, the parameters this thesis had to fix by fiat — decay rates, forcing gains, bloom coupling, the palette itself — become things a rower or a coach can move, down to writing a shader outright. Those are different products with different risks: a chooser is usable immediately and cannot be tuned into incoherence, while an editable substrate is powerful and can be. The interesting work is in the middle, and it is empirical rather than architectural. Which parameters reward being exposed, which are load-bearing enough that moving them breaks the coupling Theme 1 depends on, and whether a rower who tunes their own display stays engaged past the point where a fixed mapping is solved — that last question connects directly to the exhaustible-mapping finding, because a mapping the user can change may not be exhaustible in the same way.

Showing and hiding the numbers. P10's request for the PM5 data on a control the rower operates (Section 6.1.4) describes a testable design in Buxton's terms: a display supporting explicit user-controlled movement between a persistent low-bandwidth channel and a bursty high-bandwidth one (25). That is the traffic condition Weiser and Brown regard as definitive, made into an interface control rather than left to chance.

Objects in the flow. One extension has already been prototyped and is worth reporting because it changes what the display is for rather than how well it performs. A branch of the repository, `TouchTable`, runs the same simulation on an interactive table. Objects placed on the surface are tracked and streamed to the renderer over UDP as an identifier, a position and an orientation, each entering the simulation as a rotatable square proxy rather than as

a traced outline, and the fluid parts around them. Nothing in the numerical method had to change to support this. The solver already carries an obstacle mask — it is the first stage of the per-frame graph (Figure 3.7) and enters the pressure projection as a boundary condition on divergence (Appendix A) — because the boat cutout is itself an obstacle, and a tracked object marks that mask *dirty* in exactly the way a resize does — that is, it flags the mask as stale so the next frame recomputes it, rather than the mask being rebuilt continuously. What the table adds is not fluid capability but a way to author obstacles by hand.

Bringing that to the floor projection is a sensing problem rather than a rendering one. The ergometer installation currently senses the rower and nothing else, so the projection has no way of knowing when an object or a person enters the flow. Extending the sensed area across the projected region would let anything placed there become part of the simulation: an object set down at the edge would divert the water around it, and someone stepping in would leave a wake behind them. This is the property the surface has and the room does not — a floor that can be given rules, on which an ordinary object acquires hydrodynamic consequences it could never have on a real river. Two routes are open, and they differ in what they cost. Depth or overhead sensing across the projected area would preserve the projection setup and add a registration and calibration problem of the kind that projection-based floor displays are known to require (135). A floor display would remove the projection geometry entirely and with it the occlusion of a rower's own shadow, at the cost of the architectural quality that made projection attractive in the first place (64, 130).

The reason to want either is not that obstacle interaction is novel in itself. It is that this thesis found engagement resting on the bare fact of being answered, and every account of that in Chapter 5 concerns the rower's own movement. Whether a display that also answers to the room — to objects, to other people — deepens that coupling or dilutes it by giving the rower something else to attend to is an open question, and one this design is now close enough to test.

Comparative and technical work. A controlled comparison against the standard console, or against an explicit coaching display such as ARrow (62), would test whether ambient feedback substitutes for or complements prescriptive feedback. Measuring the remaining display-side latency along the lines in Section 3.11 would close what is left of that confound. Future evaluations could add a workload measure (54), the Intrinsic Motivation Inventory for the self-determination constructs discussed in Chapter 2 (100), or a sustained-attention measure such as the SART (98). The same architecture could support other rhythmic exercise domains where continuous peripheral feedback may be preferable to focal numerical display.

Instruments that can hold an experience

The largest open question this study raises is not about RowSim. It is about whether the instruments used to evaluate systems like it can carry what they are asked to carry, and the evidence for the question is a single item.

This thesis followed standard HCI practice. It took a published engagement scale with established reliability, adapted the wording to the rowing context as the instrument's instructions permit, and reported subscale and item-level results. That is the correct procedure, and it produced one number that will not sit still: "I lost myself in this experience," SD 1.44, the highest variance of any item administered, on a scale where nothing else exceeded 1.17. Some participants asked what "lost" was meant to signify — whether the scale wanted them to affirm it or deny it (Section 5.2.2). The word carries a positive technical sense in the absorption literature and a neutral-to-negative one in ordinary use, and the item does not disambiguate.

The scale's own statistics say the same thing more plainly. The two items making up focused attention correlate at $r = .02$ across the seventeen participants, giving that subscale a Cronbach's α of .04 where the other three subscales run from .64 to .90 (Appendix B.3). Whatever "I lost myself in this experience" measured, it was not what "I was absorbed in the activity" measured, in a sample that answered both about the same ten minutes.

That is a small drafting problem, and behind it sits a larger one about what a rating scale can carry.

The general form of the difficulty is well documented. An exhaustive description of a thing is not acquaintance with it: a complete map of a city is accurate, useful, and no substitute for having walked there. Nagel gives the philosophical statement — a complete objective account of an organism's physiology leaves untouched what it is like to be that organism, because the subjective character of experience is not the kind of thing a third-person description is built to contain (87) — and Polanyi the compact one, most relevant to a questionnaire: we can know more than we can tell (92). Where the thing to be reported exceeds the vocabulary available for reporting it, what comes back is not a smaller version of the experience but a different kind of object.

Computer science has its own discipline for this problem, and it is instructive that the discipline is not resignation. It is *abstraction*. Dijkstra put the point in terms that bear directly on this thesis: "the purpose of abstracting is *not* to be vague, but to create a new semantic level in which one can be absolutely precise" (36). An abstraction discards detail deliberately in order to become exact about something else. A sorting interface that hides

whether the implementation is quicksort or mergesort is not being imprecise; it is being precise at the level of *sorted*, which is the level the caller needs.

RowSim's display is an abstraction in exactly Dijkstra's sense, and describing it that way is more accurate than calling it ambiguous. It discards the digit and becomes precise about something else — rhythm, continuity, and the relation between a stroke and its consequence. Participants read those properties reliably and at almost no interpretive cost (Section 5.6); what they could not read, because it was deliberately discarded, was watts. The display is vague about power and exact about flow. It occupies a different semantic level, not a blurrier version of the same one.

Blackwell's history of metaphor in interface design bears on what happens to such abstractions in use. His observation is that designers tend to treat a metaphor as a mental model to be installed, whereas "metaphor users do not apply the set theories and database relations of computational knowledge structures. Instead, they treat metaphorical meaning as fuzzy, mutable, dynamic, and social" (14). Meanings drift, and a figure used often enough stops being read as a figure at all — the reason a floppy disk still means *save* to people who have never handled one. That is the fate a display like RowSim's would need in order to last. Its magenta flow currently borrows plausibility from water; a mature version would mean *this is how you are moving* without needing the water to be recognisable. Nothing in a ten-minute session can show whether that transfer happens, but it names what a longitudinal study would look for.

The measurement problem follows from the same picture. RowSim's central commitment is that a display can be exact at a level other than the numeric one. The instrument used to evaluate it, however, works only at the level of nameable propositions: twelve declarative statements, each to be agreed with on a five-point scale, each assuming that the state in question can be named accurately enough for agreement to mean the same thing to everyone. That assumption holds for most of the items — "I found the system confusing" returned an SD of 0.71 — and visibly fails for the one asking about self-loss. The seam shows exactly where the abstraction the artefact trades in has no corresponding abstraction in the instrument.

The assumption underneath this has been named before, though about design rather than about measurement. Dunne and Raby write that "dark, complex emotions are usually ignored in design," and that "nearly every other area of culture accepts that people are complicated, contradictory, and even neurotic, but not design. We view people as obedient and predictable users and consumers" (42). The point transfers cleanly, because a five-point agreement scale is that same view in instrument form: it asks a person to be legible on demand, in a vocabulary chosen in advance by someone else, about a state they may have

no settled name for. What the scale then records is treated as the experience rather than as one report of it. The artefact this thesis built declines to prescribe what the rower should attend to; the instrument used to evaluate it prescribes exactly how the rower may answer. That asymmetry is not fatal, and it is not unusual, but it is worth seeing plainly.

This is not an argument for abandoning validated scales, and nothing in these results would be improved by discarding them. The UES-SF — here in the adapted form described in Section 4.4 — did its work: it established that engagement was high, that it was not distributed by rowing ability, and that focused attention behaved differently from the other three dimensions. That last finding is the instrument revealing its own edge, which is what a good instrument does. The argument is that the edge is worth treating as a research problem rather than reporting as a limitation.

What that research might look like is an open question, and a genuinely interdisciplinary one. Experience-centred design in HCI has argued for two decades that felt experience is not reducible to task outcomes and requires methods built for it (80), and somaesthetic interaction design has developed both a method and a first-person stance for it — Höök gives a chapter to soma design methods and another to six first-person design encounters (61). The specific gap this study exposes is narrower and more tractable than the general problem: engagement instruments assume that a participant and a researcher share the valence of a word like *lost*, and they collect no evidence that this is true. Three directions seem worth pursuing. Items could be paired with a valence check, asking not only how strongly a state was experienced but whether the participant regarded it as good — which would have separated the two readings of this item at negligible cost. Instruments could allow oblique response, letting participants select or rank images, motions or metaphors rather than assent to sentences, which is the allegorical strategy applied to measurement rather than to expression. And the interview, which in this study repeatedly caught what the questionnaire missed, could be treated as the primary instrument for experiential constructs and the scale as the secondary one, rather than the reverse.

The claim is small and, I think, defensible: a field that designs systems on the premise that meaning can be conveyed without being spelled out should be curious about whether its measurement practices rest on the opposite premise. This study found its evidence for the question in a single standard deviation.

6.5 Design Considerations

Five considerations follow. Each rests on seventeen participants, one prototype, one palette, one laboratory and a single exposure, so each is stated as an observation, an interpretation, the boundary of that interpretation, and what would overturn it.

1. Make the action–effect coupling immediate and legible; other qualities are more negotiable. The most consistent source of engagement was direct correspondence between a stroke and a visible consequence (Theme 1), reported by all seventeen participants including those who disliked the palette, wanted sound, or found individual elements opaque. *Boundary:* first exposure, one visual treatment; this does not establish that coupling remains sufficient once novelty passes. *Falsifiable by:* a longitudinal study holding coupling constant while varying palette and completeness.

2. Leave the mapping open, but expect openness to be spent rather than sustained. Ambiguity engaged participants by inviting them to reverse-engineer the mapping (Theme 2) at almost no interpretive cost. But participants who solved it anticipated boredom (Theme 6). Treat an open mapping as a resource consumed by exploration, and plan for what happens once it is understood. *Boundary:* decay is anticipated within a single session, not observed across sessions. *Falsifiable by:* a multi-session study showing engagement sustained with no change to the mapping.

3. Model the loss of momentum, but also let the user command its arrest. RowSim's flow decays when rowing stops, which reads correctly as a hull losing way. What it lacks is a deliberate braking action, and participants at both ends of the expertise range asked for one (Theme 5) — a first-time rower wanting to feel agency over the boat, and a rower of thirty years naming the real manoeuvre, holding water. Where a display models a physical system, consider which of that system's states are reached passively and which are reached by doing something, and afford the second kind as actions. *Boundary:* three participants, one of whom wanted the opposite default. *Falsifiable by:* implementing a commandable arrest and finding it unused or unremarked.

4. Choose the visual register deliberately, and let it be unmistakable. The magenta palette was a deliberate refusal of realism (Section 4.1), and participants divided instructively. Eight read it against a water referent and found it wanting; one identified the rationale precisely and valued it, reporting that the distance from realism "made it clear that I was here playing in XR game rather than in reality" (P11). The lesson is not that the choice was wrong but that a display whose *behaviour* is realistic enough to summon a referent must make its departure from that referent legible as a choice, or the departure will be read as failure. *Boundary:* one palette, one lighting condition, one projection surface, which

may itself have driven the legibility complaints P8 and P13 raised. *Falsifiable by:* testing matched abstract and naturalistic palettes and finding no difference in reported coherence.

5. Decide whether the goal is engagement or training, and instrument accordingly. An ambient display rendering the effect of a movement gives an external focus of attention, which the motor-learning literature supports (133, 134). But the strongest experiential effect participants reported went further: displacement of attention away from exertion altogether (Theme 3). Those are different outcomes with different success criteria, and neither is measured by an engagement questionnaire. Ambient sports displays should state which they optimize for — workload and sustained-attention measures where the goal is training (54, 98), engagement and adherence measures where it is recreational participation. RowSim has been evaluated against the second. *Boundary:* the displacement is self-reported, not measured. *Falsifiable by:* a workload instrument showing no difference in attentional allocation between ambient and numeric conditions.

A note on evaluation instruments. Beyond RowSim, the divergence between the two focused-attention items suggests that engagement instruments applied to ambient exercise systems should distinguish immersive absorption from analytic exploration. Both are engaged states; only the first is captured by conventional absorption items, and a system that provokes deliberate experimentation will be scored as less engaging than it is.

Chapter 7

Conclusion

This thesis began from an observation about the machine before any question about how to display things on it: indoor rowing preserves the effort of locomotion while deleting its perceptual accompaniment, and fills the resulting vacancy with digits. What RowSim offers is still an interface — but one built on a different premise about where information should sit and how precisely it should speak. It occupies the emptied space continuously rather than intermittently, in the lower periphery rather than in central vision, and without resolving into any quantity that can be read off.

The two questions set out in Section 1.2 can be answered directly before the findings are drawn out of them.

RQ1 — how rowers experience and engage with an ambient mixed-reality rowing display. They engaged with it highly and uniformly, and they located its value in being answered rather than in being informed. Three of the four adapted UES-SF subscales sat near the ceiling and none differed between casual and elite rowers (Section 5.4); all seventeen described the correspondence between what they did and what the floor showed, which is the most widely expressed theme in the corpus (Section 5.6); and the under-specified mapping produced exploration rather than confusion, the rhythm-disruption item returning the lowest mean and variance of any item in the instrument. The fourth subscale, focused attention, is the exception and the more interesting reading: at 3.82 it alone stays off the ceiling, and Section 5.2.2 shows the reason is a split within it — participants reported absorption in the activity while declining to say they lost themselves in it. Two qualifications belong with that answer. The engagement was recorded in a single ten-minute session with no comparison condition, so it characterises experience *during* RowSim rather than relative to a console (Section 4.7); and the mechanism was mostly calibration, which terminates, with five participants predicting their own boredom from inside the session.

RQ2 — how rowing performance during RowSim differs between casual and elite rowers. Decisively on the physical measures and not at all on the experiential ones. Elite rowers applied 2.6 times the peak drive force, produced 3.7 times the mean power and did 4.3 times the work per stroke, and all seven force-and-output comparisons survive Holm

correction; no timing measure and no engagement measure approaches significance either before or after (Section 5.4). Because the same test on the same two groups separates them on one family and not the other, the null on engagement is informative rather than merely underpowered — though the elite arm has five participants, which bounds how small an engagement difference this design could have detected.

Five things were learned, and they are worth stating as what they change rather than as what was measured.

Responsiveness is not a scarce resource. The study set out expecting expertise to matter, and on the rowing it did: trained and untrained participants separated on every measure of force and output, several of them decisively. On engagement they did not separate at all. The reason the interviews give is that what participants valued was not achievement rendered visible but the bare fact of being answered — and a room that responds to an eight-watt stroke responds just as promptly to a hundred-and-sixty-watt one. A display whose appeal rests on coupling rather than on scoring has no obvious reason to reward one level of performance over another, and nothing in this sample suggests it did, which would be an unusual property for a piece of sports technology and a useful one if the goal is participation rather than ranking.

Calm design survives a working body. Weiser and Brown drew a boundary around their own programme, doubting it would extend to activities where the point is to be excited, and ambient-display research has largely respected that boundary by staying in offices and homes. Moving the competing demand from cognition to physical exertion did not break the peripheral channel. It carried a ten-minute session alone, with the ergometer's own console covered and no clock visible, and the participant best placed to judge described the result in rowing coaching's own words for peripheral attention. On the evidence here the boundary looks to have been drawn more conservatively than it needed to be — though peripheral uptake was reported rather than measured, and Section 6.4 sets out the experiment that would settle it.

Interpretive openness is spent, not held. Participants' responses were consistent with what the design literature predicts of ambiguity: it appeared to invite engagement rather than to produce frustration. The display was not varied in ambiguity, so this is a pattern consistent with the prediction rather than a test of it. But the mechanism was mostly calibration rather than meaning-making, and calibration finishes. Participants told us so from inside a single session, and hedonic adaptation predicts the rest (45). An open mapping should be budgeted like a consumable, with a plan for the state after it has been solved — which is a different design problem from the one this thesis set out to address, and a harder one.

Restoring the missing motion is a coupling problem, not a rendering one. The thesis began by treating the ergometer as a perceptual anomaly rather than as a machine short of a screen: locomotor effort with an optic array that does not change. Filling that array turned out to be the easy half. Two participants at opposite ends of the expertise range independently reported that the flow did not read as theirs, and both located the reason in the same place — it could not be stopped. On the water, holding water is an action with an immediate consequence; in RowSim, stopping is only the absence of action, resolved by a decay the rower waits out. What specifies self-motion is not motion in the visual field but motion that answers to what the observer does, and a display that accelerates on demand while decelerating on its own schedule supplies half of that. This is the finding that could not have been arrived at from the ambient-display literature alone, because that literature concerns people watching motion they are not producing. It also happens to be cheap to act on.

A protocol that declines to prescribe turns behaviour into evidence. Setting no target and permitting rest cost the study nothing and bought it a measure it would otherwise not have had. Whether someone kept rowing became a datum rather than an instruction followed, self-reports about persistence could be checked against logs the participants could not see, and the one expert who structured his own session into pieces did something the protocol would have forbidden had it specified anything. The methodological point generalizes past rowing: where a system is designed to leave interpretation open, a study that closes it measures the wrong thing.

What the thesis cannot say is bounded by two things. Engagement was recorded once, and participants themselves expect it to fall. And the peripherality claim rests on report rather than on a demonstration that peripherally available information was taken up — a gap eye tracking would not close, since foveation is not what is at issue, and a detection paradigm would. Section 6.4 sets out both, along with a specific and buildable gap the participants identified themselves: a display that models the loss of momentum but affords no way to command it.

One last thread runs back to the beginning. The argument for RowSim is that some things are better conveyed by withholding precision than by supplying it; the instrument used to evaluate it assumes the opposite, that a felt state can be named accurately enough for agreement to mean the same thing to everyone. The item asking about the most experientially loaded state was the one that fractured, because participants could not tell whether being *lost* was something to affirm. A discipline that builds artefacts on one premise and evaluates them on its opposite is worth sitting with. Section 6.4 argues the case; the evidence here is one standard deviation and one question asked out loud.

What is offered here is not RowSim as software but an account of how ambient, interpretively open feedback behaves in front of a working body — and of what made that account possible: a trustworthy stroke record, a mapping made to be read rather than decoded, and a protocol that refused to instruct.

Bibliography

[1] Morgan Ames and Anind K. Dey. Description of design dimensions and evaluation for ambient displays. Technical Report UCB/CSD-02-1211, EECS Department, University of California, Berkeley, 2002.

[2] Roberta Antonini Philippe, Sarah Morgana Singer, Joshua E. E. Jaeger, Michele Biasutti, and Scott Sinnett. Achieving flow: An exploratory investigation of elite college athletes and musicians. *Frontiers in Psychology*, 13:831508, March 2022. doi: 10.3389/fpsyg.2022.831508.

[3] Apple Inc. The qualities of great design. WWDC18 Session 801, Apple Worldwide Developers Conference, 2018. URL `https://developer.apple.com/videos/play/wwdc2018/801/`.

[4] Apple Inc. Metal: Documentation and programming guide. Apple Developer Documentation, 2024. URL `https://developer.apple.com/documentation/metal`.

[5] Duarte Araújo, Keith Davids, and Robert Hristovski. The ecological dynamics of decision making in sport. *Psychology of Sport and Exercise*, 7(6):653–676, 2006. doi: 10.1016/j.psychsport.2006.07.002.

[6] Sebastian Arndt, Andrew Perkis, and Jan-Niklas Voigt-Antons. Using virtual reality and head-mounted displays to increase performance in rowing workouts. In *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports*, pages 45–50, Seoul, Republic of Korea, October 2018. ACM. doi: 10.1145/3265845.3265848.

[7] Simon Attfield, Gabriella Kazai, Mounia Lalmas, and Benjamin Piwowarski. Towards a science of user engagement. *Proceedings of the Workshop on Search and Exploration of Xtremely Large Collections (WSDEX)*, 2011.

[8] A. Baca and P. Kornfeind. Rapid feedback systems for elite sports training. *IEEE Pervasive Computing*, 5(4):70–76, October 2006. doi: 10.1109/mprv.2006.82.

[9] Saskia Bakker, Elise van den Hoven, and Berry Eggen. Peripheral interaction: characteristics and considerations. *Personal and Ubiquitous Computing*, 19:239–254, 2015. doi: 10.1007/s00779-014-0775-2.

[10] Benoît G. Bardy and Marielle Laurent. How is body orientation controlled during somersaulting? *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):963–977, 1998. doi: 10.1037/0096-1523.24.3.963.

[11] Artur Becker, Henrik Herrebrøden, Victor E. González Sánchez, Kristian Nymoen, Carla Maria Dal Sasso Freitas, Jim Torresen, and Alexander Refsum Jensenius. Functional data analysis of rowing technique using motion capture data. In *Proceedings of the 6th International Conference on Movement and Computing (MOCO '19)*, pages 1–8. ACM, 2019. doi: 10.1145/3347122.3347135.

[12] Nikolai A. Bernstein. *The Co-ordination and Regulation of Movements*. Pergamon Press, Oxford, UK, 1967.

[13] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters/CRC Press, 2005.

[14] Alan F. Blackwell. The reification of metaphor as a design tool. *ACM Transactions on Computer-Human Interaction*, 13(4):490–530, 2006. doi: 10.1145/1188816.1188820.

[15] Iñigo Borges, Santiago Veiga, and Pablo González-Frutos. The evaluation of physical performance in rowing ergometer: A systematic review. *Journal of Functional Morphology and Kinesiology*, 10(4):437, November 2025. doi: 10.3390/jfmk10040437.

[16] Frank Brady. Contextual interference: A meta-analytic study. *Perceptual and Motor Skills*, 99(1):116–126, 2004. doi: 10.2466/pms.99.1.116-126.

[17] Dennis M. Bramble and Daniel E. Lieberman. Endurance running and the evolution of Homo. *Nature*, 432(7015):345–352, 2004. doi: 10.1038/nature03052.

[18] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa.

[19] Virginia Braun and Victoria Clarke. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597, 2019. doi: 10.1080/2159676x.2019.1628806.

[20] Virginia Braun and Victoria Clarke. One size fits all? what counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3):328–352, 2021. doi: 10.1080/14780887.2020.1769238.

[21] Virginia Braun and Victoria Clarke. To saturate or not to saturate? questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, 13(2):201–216, 2021. doi: 10.1080/2159676X.2019.1704846.

[22] Robert Bridson. *Fluid Simulation for Computer Graphics*. CRC Press, 2nd edition, 2015.

[23] John Brooke. SUS: a 'quick and dirty' usability scale. In Patrick W. Jordan, Bruce Thomas, Bernard A. Weerdmeester, and Ian L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis, 1996. doi: 10.1201/9781498710411-35.

[24] PJ Burns. *The use of wearable ambient displays to enhance awareness of physical activity*. PhD thesis, University of Tasmania, 2016.

[25] William Buxton. Integrating the periphery and context: A new model of telematics. In *Proceedings of Graphics Interface '95*, pages 239–246, 1995.

[26] Kelly Caine. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016. doi: 10.1145/2858036.2858498.

[27] Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. Tri-council policy statement: Ethical conduct for research involving humans – TCPS 2 (2022). Panel on Research Ethics, Government of Canada, December 2022. URL `https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html`.

[28] Jia Yi Chow, Keith Davids, Chris Button, and Ian Renshaw. *Nonlinear Pedagogy: An Effective Approach to Support Skill Acquisition*. Routledge, 2016.

[29] William S. Cleveland and Robert McGill. Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984. doi: 10.1080/01621459.1984.10478080.

[30] Jacob Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. doi: 10.1037/0033-2909.112.1.155.

[31] Keenan Crane, Ignacio Llamas, and Sarah Tariq. Real-time simulation and rendering of 3D fluids. In Hubert Nguyen, editor, *GPU Gems 3*, chapter 30. Addison-Wesley, 2007.

[32] Maria I. Cruz, Hugo Sarmento, Ana M. Amaro, Luís Roseiro, and Beatriz B. Gomes. Advancements in performance monitoring: A systematic review of sensor technologies in rowing and canoeing biomechanics. *Sports*, 12(9):254, September 2024. doi: 10.3390/sports12090254.

[33] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper & Row, 1990.

[34] Andrew Dahley, Craig Wisneski, and Hiroshi Ishii. Water lamp and pinwheels: Ambient projection of digital information into architectural space. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, pages 269–270, Los Angeles, CA, USA, 1998. ACM. doi: 10.1145/286498.286750.

[35] Ed Diener, Richard E. Lucas, and Christie Napa Scollon. Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist*, 61 (4):305–314, 2006. doi: 10.1037/0003-066X.61.4.305.

[36] Edsger W. Dijkstra. The humble programmer. *Communications of the ACM*, 15(10): 859–866, 1972. doi: 10.1145/355604.361591. ACM Turing Award Lecture.

[37] Paul A. M. Dirac. The evolution of the physicist's picture of nature. *Scientific American*, 208(5):45–53, 1963. doi: 10.1038/scientificamerican0563-45.

[38] Martin Dobiasch, Björn Krenn, Robert P. Lamberts, and Arnold Baca. The effects of visual feedback on performance in heart rate- and power-based-tasks during a constant load cycling test. *Journal of Sports Science and Medicine*, 21(1):49–57, 2022. doi: 10.52082/jssm.2022.49.

[39] Pavel Dobryakov. WebGL fluid simulation. Open-source project, MIT licence, 2019. URL `https://github.com/PavelDoGreat/WebGL-Fluid-Simulation`.

[40] Paul Dourish. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, 2004.

[41] Douglas Dow, Ryan Andrews, Alejandra Garcia, Brandon Dryer, and Scott Bonney. Rowing training system for on-the-water rehabilitation and sport. In *Proceedings of the 8th International Conference on Body Area Networks (BodyNets '13)*, 2013. doi: 10.4108/icst.bodynets.2013.253697.

[42] Anthony Dunne and Fiona Raby. *Speculative Everything: Design, Fiction, and Social Dreaming*. MIT Press, Cambridge, MA, 2013.

[43] Ronald Fedkiw, Jos Stam, and Henrik Wann Jensen. Visual simulation of smoke. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, 2001. doi: 10.1145/383259.383260.

[44] Patrick Tobias Fischer, Christian Zöllner, Thilo Hoffmann, Sebastian Piatza, and Eva Hornecker. Beyond information and utility: Transforming public spaces with media façades. *IEEE Computer Graphics and Applications*, 33(2):38–46, March 2013. doi: 10.1109/mcg.2012.126.

[45] Shane Frederick and George Loewenstein. Hedonic adaptation. In Daniel Kahneman, Ed Diener, and Norbert Schwarz, editors, *Well-Being: The Foundations of Hedonic Psychology*, pages 302–329. Russell Sage Foundation, New York, NY, 1999.

[46] Shaun Gallagher. *How the Body Shapes the Mind*. Oxford University Press, 2005.

[47] William W. Gaver, Jacob Beaver, and Steve Benford. Ambiguity as a resource for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 233–240, Ft. Lauderdale, FL, USA, 2003. ACM. doi: 10.1145/642611.642653.

[48] James J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

[49] Franz Gravenhorst, Christoph Thiem, Bernd Tessendorf, Rolf Adelsberger, Bert Arnrich, and Gerhard Tröster. Sonicseat: A seat position tracker based on ultrasonic sound measurements for rowing technique analysis. In *Proceedings of the 7th International Conference on Body Area Networks*, Oslo, Norway, 2012. ACM. doi: 10.4108/icst.bodynets.2012.249917.

[50] Mark A. Guadagnoli and Timothy D. Lee. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, 36(2):212–224, 2004. doi: 10.3200/jmbr.36.2.212-224.

[51] Michelle C. Haas, Larissa Wild, Leander Schneeberger, Eveline S. Graf, and Anna Lisa Martin-Niedecken. Exploring mixed-reality exergames for sports rehabilitation: Design insights and evaluation findings. *JMIR Serious Games*, 13: e68431, October 2025. doi: 10.2196/68431.

[52] Mark J. Harris. Fast fluid dynamics simulation on the GPU. In Randima Fernando, editor, *GPU Gems: Programming Techniques, Tips, and Tricks for Real-Time Graphics*, chapter 38. Addison-Wesley, 2004.

[53] Sandra G. Hart. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909.

[54] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, Amsterdam, Netherlands, 1988. doi: 10.1016/s0166-4115(08)62386-9.

[55] Marc Hassenzahl. User experience (UX): towards an experiential perspective on product quality. In *Proceedings of the 20th Conference on l'Interaction Homme-Machine (IHM '08)*, pages 11–15. ACM, 2008. doi: 10.1145/1512714.1512717.

[56] Marc Hassenzahl and Noam Tractinsky. User experience - a research agenda. *Behaviour & Information Technology*, 25(2):91–97, 2006. doi: 10.1080/01449290500330331.

[57] Marc Hassenzahl, Michael Burmester, and Franz Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In Gerd Szwillus and Jürgen Ziegler, editors, *Mensch & Computer 2003*, pages 187–196, Wiesbaden, Germany, 2003. Vieweg+Teubner. doi: 10.1007/978-3-322-80058-9_19.

[58] Xiaohua He and Li Wei. Real-time feedback enhances motor learning and motivation in youth team sports through augmented reality tools. *Frontiers in Psychology*, 16: 1661936, December 2025. doi: 10.3389/fpsyg.2025.1661936.

[59] Martin Hedlund, Cristian Bogdan, Gerrit Meixner, and Andrii Matviienko. Rowing beyond: Investigating steering methods for rowing-based locomotion in virtual environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*, pages 1–17. ACM, 2024. doi: 10.1145/3613904.3642192.

[60] R. Hohmuth, D. Schwensow, H. Malberg, and M. Schmidt. A wireless rowing measurement system for improving the rowing performance of athletes. *Sensors*, 23 (3):1060, 2023. doi: 10.3390/s23031060.

[61] Kristina Höök. *Designing with the Body: Somaesthetic Interaction Design*. MIT Press, 2018.

[62] Elena Iannucci, Zhutian Chen, Iro Armeni, Marc Pollefeys, Hanspeter Pfister, and Johanna Beyer. ARrow: A real-time AR rowing coach. In *EuroVis 2023 – Short Papers*. The Eurographics Association, 2023. doi: 10.2312/evs.20231046.

[63] Katherine Isbister, Kristina Höök, Jarmo Laaksolahti, and Michael Sharp. The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect. *International Journal of Human-Computer Studies*, 65(4):315–328, 2007. doi: 10.1016/j.ijhcs.2006.11.017.

[64] Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 234–241, Atlanta, GA, USA, 1997. ACM. doi: 10.1145/258549.258715.

[65] Judith Jimenez-Diaz, Karla Chaves-Castro, and Maria Morera-Castro. Effect of self-controlled and regulated feedback on motor skill performance and learning: A meta-analytic study. *Journal of Motor Behavior*, 53(3):385–398, 2021. doi: 10.1080/00222895.2020.1782825.

[66] Scott R. Klemmer, Björn Hartmann, and Leila Takayama. How bodies matter: five themes for interaction design. In *Proceedings of the 6th Conference on Designing Interactive Systems (DIS '06)*, pages 140–149. ACM, 2006. doi: 10.1145/1142405.1142429.

[67] Valery Kleshnev. *The Biomechanics of Rowing*. The Crowood Press, 2016.

[68] Kristina Knaving, Paweł Wozniak, Morten Fjeld, and Staffan Björk. Flow is not enough: Understanding the needs of advanced amateur runners to design motivation technology. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, pages 2013–2022. ACM, 2015. doi: 10.1145/2702123.2702542.

[69] Daniel Lakens. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in Psychology*, 4:863, 2013. doi: 10.3389/fpsyg.2013.00863.

[70] Markus Lappe, Frank Bremmer, and A. V. van den Berg. Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3(9):329–336, 1999. doi: 10.1016/S1364-6613(99)01364-9.

[71] Bettina Laugwitz, Theo Held, and Martin Schrepp. Construction and evaluation of a user experience questionnaire. In Andreas Holzinger, editor, *HCI and Usability for Education and Work (USAB 2008)*, volume 5298 of *Lecture Notes in Computer Science*, pages 63–76, Berlin, Heidelberg, 2008. Springer. doi: 10.1007/978-3-540-89350-9_6.

[72] Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P. O. S. Vermeeren, and Joke Kort. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, pages 719–728. ACM, 2009. doi: 10.1145/1518701.1518813.

[73] Xuecheng Li, Zhengyu Wu, and Ting Han. Gamification-based VR rowing simulation system. In Masaaki Kurosu, editor, *Human-Computer Interaction. Recognition and Interaction Technologies*, volume 11567 of *Lecture Notes in Computer Science*, pages 484–493, Cham, 2019. Springer International Publishing. doi: 10.1007/978-3-030-22643-5_38.

[74] Christoph Lürig, Tilo Mentler, and Sven Karstens. Row your boat in VR and solve thinking exercises on the way: The brain-row challenge. In *30th ACM Symposium on Virtual Reality Software and Technology*, pages 1–2, Trier, Germany, October 2024. doi: 10.1145/3641825.3689498.

[75] Kirsti Malterud, Volkert Dirk Siersma, and Ann Dorrit Guassora. Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13):1753–1760, 2016. doi: 10.1177/1049732315617444.

[76] Jennifer Mankoff, Anind K. Dey, Gary Hsieh, Julie Kientz, Scott Lederer, and Morgan Ames. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 169–176, Ft. Lauderdale, FL, USA, 2003. ACM. doi: 10.1145/642611.642642.

[77] Derek T. Y. Mann, A. Mark Williams, Paul Ward, and Christopher M. Janelle. Perceptual-cognitive expertise in sport: a meta-analysis. *Journal of Sport and Exercise Psychology*, 29(4):457–478, 2007. doi: 10.1123/jsep.29.4.457.

[78] Joe Marshall and Steve Benford. Using fast interaction to create intense experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 1255–1264. ACM, 2011. doi: 10.1145/1978942.1979129.

[79] Aleka McAdams, Eftychios Sifakis, and Joseph Teran. A parallel multigrid poisson solver for fluids simulation on large grids. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '10)*, pages 65–73. Eurographics Association, 2010. doi: 10.2312/SCA/SCA10/065-073.

[80] John McCarthy and Peter Wright. *Technology as Experience*. MIT Press, Cambridge, MA, 2004.

[81] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, 1994.

[82] Florian Mueller, Frank Vetere, Martin R. Gibbs, Darren Edge, Stefan Agamanolis, and Jennifer G. Sheridan. Jogging over a distance between europe and australia. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*, pages 189–198. ACM, 2010. doi: 10.1145/1866029.1866062.

[83] Florian Mueller, Darren Edge, Frank Vetere, Martin R. Gibbs, Stefan Agamanolis, Bert Bongers, and Jennifer G. Sheridan. Designing sports: a framework for exertion games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, pages 2651–2660. ACM, 2011. doi: 10.1145/1978942.1979330.

[84] Florian 'Floyd' Mueller and Damon Young. Five lenses for designing exertion experiences. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*, pages 2473–2487. ACM, 2017. doi: 10.1145/3025453.3025746.

[85] Tamara Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[86] Elizabeth L. Murnane, Xin Jiang, Anna Kong, Michelle Park, Weili Shi, Connor Soohoo, Luke Vink, Iris Xia, Xin Yu, John Yang-Sammataro, Grace Young, Jenny Zhi, Paula Moya, and James A. Landay. Designing ambient narrative-based interfaces to reflect and motivate physical activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Honolulu, HI, USA, 2020. ACM. doi: 10.1145/3313831.3376478.

[87] Thomas Nagel. What is it like to be a bat? *The Philosophical Review*, 83(4): 435–450, 1974. doi: 10.2307/2183914.

[88] Heather L. O'Brien and Elaine G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, 2010. doi: 10.1002/asi.21229.

[89] Heather L. O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112:28–39, 2018. doi: 10.1016/j.ijhcs.2018.01.004.

[90] Chris Parnin and Carsten Görg. Design guidelines for ambient software visualization in the workplace. In *2007 4th IEEE International Workshop on Visualizing Software for Understanding and Analysis*, pages 18–25, Banff, AB, Canada, 2007. IEEE. doi: 10.1109/vissof.2007.4290695.

[91] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale. State of the art of sports data visualization. *Computer Graphics Forum*, 37(3):663–686, June 2018. doi: 10.1111/cgf.13447.

[92] Michael Polanyi. *The Tacit Dimension*. Doubleday, Garden City, NY, 1966.

[93] Dees Postma, Dennis Reidsma, Robby Van Delden, and Armağan Karahanoğlu. From metrics to experiences: Investigating how sport data shapes the social context, self-determination and motivation of athletes. *Interacting with Computers*, 38(3): 394–408, 2026. doi: 10.1093/iwc/iwae012.

[94] Zachary Pousman and John Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 67–74, 2006. doi: 10.1145/1133265.1133277.

[95] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *Proceedings of SIGGRAPH '98*, pages 179–188. ACM, 1998. doi: 10.1145/280814.280861.

[96] Johan Redström, Tobias Skog, and Lars Hallnäs. Informative art: Using amplified artworks as information displays. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, pages 103–114, Elsinore, Denmark, 2000. ACM. doi: 10.1145/354666.354677.

[97] Ian Renshaw, Keith Davids, Daniel Newcombe, and Will Roberts. *The Constraints-Led Approach: Principles for Sports Coaching and Practice Design*. Routledge, 2019.

[98] Ian H. Robertson, Tom Manly, Jackie Andrade, Bart T. Baddeley, and Jenny Yiend. 'oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6):747–758, 1997. doi: 10.1016/s0028-3932(97)00015-8. Introduces the Sustained Attention to Response Task (SART).

[99] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pages 261–270, Copenhagen, Denmark, 2010. ACM. doi: 10.1145/1864349.1864372.

[100] Richard M. Ryan. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3): 450–461, 1982. doi: 10.1037/0022-3514.43.3.450. Foundational work on intrinsic motivation and the distinction between informational and controlling feedback.

[101] Richard M. Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78, 2000. doi: 10.1037/0003-066x.55.1.68.

[102] Alan W. Salmoni, Richard A. Schmidt, and Charles B. Walter. Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin*, 95(3): 355–386, 1984. doi: 10.1037/0033-2909.95.3.355.

[103] Nina Schaffert, Klaus Mattes, and Alfred O. Effenberg. Examining effects of acoustic feedback on perception and modification of movement patterns in on-water rowing training. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, pages 122–129, Coimbra, Portugal, September 2011. doi: 10.1145/2095667.2095685.

[104] Richard A. Schmidt. A schema theory of discrete motor skill learning. *Psychological Review*, 82:225–260, 1975. doi: 10.1037/h0076770.

[105] Jaehyuk Shim, Youngnoh Goh, Hyejin Kim, Seungchan Lim, and Daseong Han. A rowing game based on motion similarity of two players. In *Proceedings of the 31st International Conference on Computer Animation and Social Agents*, pages 83–85, Beijing, China, May 2018. doi: 10.1145/3205326.3205362.

[106] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, 20(1):21–53, 2013. doi: 10.3758/s13423-012-0333-8.

[107] Tobias Skog, Sara Ljungblad, and Lars Erik Holmquist. Between aesthetics and utility: Designing ambient information visualizations. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis 2003)*, pages 233–240. IEEE, 2003. doi: 10.1109/infvis.2003.1249031.

[108] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138.

[109] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997. doi: 10.1162/pres.1997.6.6.603.

[110] Richard M. Smith and Warwick L. Spinks. Discriminant analysis of biomechanical differences between novice, good and elite rowers. *Journal of Sports Sciences*, 13(5): 377–385, 1995. doi: 10.1080/02640419508732253.

[111] Clare Soper and Patria A. Hume. Towards an ideal rowing technique for performance: the contributions of biomechanics. *Sports Medicine*, 34(12):825–848, 2004. doi: 10.2165/00007256-200434120-00003.

[112] Joanna Sorysz, Meagan B. Loerakker, and Bettina Eska. Feedback solutions in rowing training. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '24 Adjunct)*, pages 450–453. ACM, 2024. doi: 10.1145/3675094.3678500.

[113] Jos Stam. Stable fluids. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '99)*, pages 121–128. ACM, 1999. doi: 10.1145/311535.311548.

[114] Jos Stam. Real-time fluid dynamics for games. In *Proceedings of the Game Developer Conference (GDC)*, 2003.

[115] Martyn Standage and Richard M. Ryan. Self-determination theory in sport and exercise. In Gershon Tenenbaum and Robert C. Eklund, editors, *Handbook of Sport Psychology*, pages 37–56. Wiley, 4th edition, 2020. doi: 10.1002/9781119568124.ch3.

[116] John Steinhoff and David Underhill. Modification of the Euler equations for "vorticity confinement": Application to the computation of interacting vortex rings. *Physics of Fluids*, 6(8):2738–2744, 1994. doi: 10.1063/1.868164.

[117] Norbert A. Streitz, Shin'ichi Konomi, and Heinz-Jürgen Burkhardt, editors. *Cooperative Buildings: Integrating Information, Organization, and Architecture*, volume 1370 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 1998. doi: 10.1007/3-540-69706-3.

[118] Lucy A. Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge, UK, 1987.

[119] Christian Swann, Lee Crust, and Stewart A. Vella. New directions in the psychology of optimal performance in sport: flow and clutch states. *Current Opinion in Psychology*, 16:48–53, August 2017. doi: 10.1016/j.copsyc.2017.03.032.

[120] Issey Takahashi, Mika Oki, Baptiste Bourreau, Itaru Kitahara, and Kenji Suzuki. FUTUREGYM: A gymnasium with interactive floor projection for children with special needs. *International Journal of Child-Computer Interaction*, 15:37–47, March 2018. doi: 10.1016/j.ijcci.2017.12.002.

[121] Javier Vales-Alonso, Pablo López-Matencio, Francisco J. González-Castaño, Honorio Navarro-Hellín, Pedro J. Baños-Guirao, Francisco J. Pérez-Martínez, Rafael P. Martínez-Alvarez, Daniel González-Jiménez, Felipe Gil-Castiñeira, and Richard Duro-Fernández. Ambient intelligence systems for personalized sport training. *Sensors*, 10(3):2359–2385, March 2010. doi: 10.3390/s100302359.

[122] Robert J. Vallerand. Intrinsic and extrinsic motivation in sport and physical activity: A review and a look at the future. In Gershon Tenenbaum and Robert C. Eklund, editors, *Handbook of Sport Psychology*, pages 59–83. Wiley, 3rd edition, 2007. doi: 10.1002/9781118270011.ch3.

[123] Robby van Delden, Sascha Bergsma, Koen Vogel, Dees Postma, Randy Klaassen, and Dennis Reidsma. VR4VRT: Virtual reality for virtual rowing training. In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '20)*, pages 388–392. ACM, 2020. doi: 10.1145/3383668.3419865.

[124] Joachim Von Zitzewitz, Peter Wolf, Vladimir Novaković, Mathias Wellner, Georg Rauter, Andreas Brunschweiler, and Robert Riener. Real-time rowing simulator with multimodal feedback. *Sports Technology*, 1(6):257–266, January 2008. doi: 10.1002/jst.65.

[125] Zixuan Wang and Burkhard Claus Wünsche. Space war: Investigating the impact of a loss aversion strategy and personality traits in a VR rowing exergame. In *Proceedings of the 2024 Australasian Computer Science Week (ACSW '24)*, pages 130–139. ACM, 2024. doi: 10.1145/3641142.3641173.

[126] William H. Warren. The dynamics of perception and action. *Psychological Review*, 113(2):358–389, 2006. doi: 10.1037/0033-295x.113.2.358.

[127] Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3): 94–104, 1991. doi: 10.1038/scientificamerican0991-94.

[128] Mark Weiser and John Seely Brown. Designing calm technology. Technical report, Xerox PARC, 1996.

[129] Carolee J. Winstein and Richard A. Schmidt. Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):677–691, 1990. doi: 10.1037/0278-7393.16.4.677.

[130] Craig Wisneski, Hiroshi Ishii, Andrew Dahley, Matt Gorbet, Scott Brave, Brygg Ullmer, and Paul Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In Norbert A. Streitz, Shin'ichi Konomi, and Heinz-Jürgen Burkhardt, editors, *Cooperative Buildings: Integrating Information, Organization, and Architecture*, volume 1370 of *Lecture Notes in Computer Science*, pages 22–32, Berlin, Heidelberg, 1998. Springer. doi: 10.1007/3-540-69706-3_4.

[131] Bob G. Witmer and Michael J. Singer. Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3): 225–240, 1998. doi: 10.1162/105474698565686.

[132] Yueyao Wu, Xiaoyu Zhao, Zhengyu Zhang, and Jiuchun Ren. Optimizing wireless signal strength on rowing boat: A 3D simulation approach to overcoming human occlusion. In *Proceedings of the 2023 9th International Conference on Communication and Information Processing (ICCIP '23)*, pages 301–307. ACM, 2023. doi: 10.1145/3638884.3638930.

[133] Gabriele Wulf. Attentional focus and motor learning: a review of 15 years. *International Review of Sport and Exercise Psychology*, 6:77–104, 2013. doi: 10.1080/1750984x.2012.723728.

[134] Gabriele Wulf and Rebecca Lewthwaite. Optimizing performance through intrinsic motivation and attention for learning: the OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, 23(5):1382–1414, 2016. doi: 10.3758/s13423-015-0999-9.

[135] Chun Xie, Hidehiko Shishido, Yoshinari Kameda, Kenji Suzuki, and Itaru Kitahara. A calibration method for large-scale projection based floor display system. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 725–726. IEEE, 2018. doi: 10.1109/VR.2018.8446433.

[136] Xiaoyu Zhao, Yueyao Wu, Zhengyu Zhang, and Jiuchun Ren. Design of intelligent indoor rowing training pool digitization system based on multisource data fusion. In *Proceedings of the 2023 9th International Conference on Communication and Information Processing (ICCIP '23)*, pages 112–118. ACM, 2023. doi: 10.1145/3638884.3638901.

[137] John Zimmerman and Jodi Forlizzi. Research through design in HCI. In Judith S. Olson and Wendy A. Kellogg, editors, *Ways of Knowing in HCI*, pages 167–189. Springer, New York, NY, 2014. doi: 10.1007/978-1-4939-0378-8_8.

[138] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2007. doi: 10.1145/1240624.1240704.

Appendix A

Extended System Implementation Detail

This appendix collects the implementation-level detail summarized in Chapter 3: the Metal resource architecture, the full numerical discretization of the fluid and wave solvers, the compute/fragment branch-divergence matrix, the PM5 telemetry service and characteristic layout, the runtime configuration surface, additional installation and output views, and the structure of the per-session logging bundle. It is provided for readers who want reconstruction-level detail; nothing here is required to follow the study in Chapters 4–6. Table A.1 collects the symbols used in the discretization that follows. Throughout, all difference operators are written in grid units, so a cell spacing of $\Delta x = 1$ is assumed and does not appear in the expressions.

A.1 Code and Data Availability

The RowSim implementation described in Chapter 3 and in this appendix is maintained in a Git repository at `https://github.com/arshshah07/RowSim`. Two branches are relevant to this thesis. The main branch is the version deployed for the study reported in Chapters 4–5; it is the code whose configuration each session's `manifest.json` records, so a given session can be matched to the software that produced it. The `TouchTable` branch is the tabletop variant discussed in Section 6.4, in which physical objects placed on an interactive surface are sensed and written into the obstacle field. It shares the solver described in Section A.3 unchanged and differs only in how obstacles are authored. The firmware for the handle-mounted inertial unit (Section A.6) is maintained separately at `https://github.com/arshshah07/ArduRowing`.

The participant data underlying Chapter 5 is held under the approved ethics protocol: the consent obtained covers analysis and reporting by the research team rather than open release, so the session bundles described in Section A.9 below are not public. The scripts that produce every figure and table from them are part of the thesis source, and the de-identified values they compute are reported in full rather than only in aggregate — per-participant

summaries in Table 5.4, item-level distributions in Figure 5.5 — so the analyses can be checked against the printed numbers even where the raw logs cannot be redistributed.

A.2 Metal Resource Architecture

All Metal objects are obtained through a single shared-device abstraction (4) that owns the `MTLDevice`, command queue, default shader library, and a pipeline-state cache keyed by function name, so repeated passes do not pay pipeline-construction cost twice. Two lightweight wrappers sit above this device abstraction:

- A *compute-shader wrapper* represents a single Metal compute kernel and derives its threadgroup width from the pipeline's own limits:

$$w_{tg} = \min\big(\,\text{execWidth},\ \text{maxThreads}\,\big), \qquad h_{tg} = 1, \tag{A.1}$$

  where execWidth is the pipeline's thread-execution width and maxThreads is its maximum total threads per threadgroup. This one-dimensional dispatch geometry is a conservative, portable choice: it favours broad device compatibility and simple dispatch arithmetic over the locality optimizations a hand-tuned two-dimensional stencil kernel might achieve.

- A *render-shader wrapper* represents fragment passes that draw a full-screen quad. The final visualization pass always uses this path; the simulation itself uses compute dispatches on Apple Silicon and fragment fallbacks otherwise.

Simulation fields are stored in *ping-pong slabs*: each slab holds two GPU-private textures (`ping`, `pong`) with shader-read, shader-write, and render-target usage. A pass reads one texture, writes the other, and swaps the pair after completion, making resource ownership explicit and avoiding undefined simultaneous read/write behaviour. All simulation fields use half-float (`rg16Float`) storage to halve bandwidth relative to full float; the drawable itself is a $4\times$ MSAA `bgra8Unorm` colour target.

Slabs are allocated exactly once, from the `MTKView` bounds at the moment the renderer is constructed, and are never reallocated. On a resize or a transition to fullscreen the textures keep their original dimensions; the final visualization pass samples them with normalized coordinates, so they stretch to whatever drawable is presented. This is deliberate — it preserves fluid state across a resize instead of discarding it — but it means the simulation grid is decoupled from both the window and the display. The installation ran fullscreen

throughout the study on two WXGA floor projectors at 1280×800 each, driven from a 1920×1080 host display. Nothing in that arrangement changed between sessions.

Because the grid size follows from a runtime bound rather than from a constant in the source, it cannot be read off the code, and it is not written to the manifest. It can, however, be read directly off a GPU capture. Figure A.1 shows the frame debugger stopped inside the final visualization fragment function with the density slab bound: `RG16Float`, $1{,}920 \times 1{,}080 \times 1$. That confirms two claims in this section at once — the half-float storage format, and a simulation grid matching the host display rather than the projector resolution.

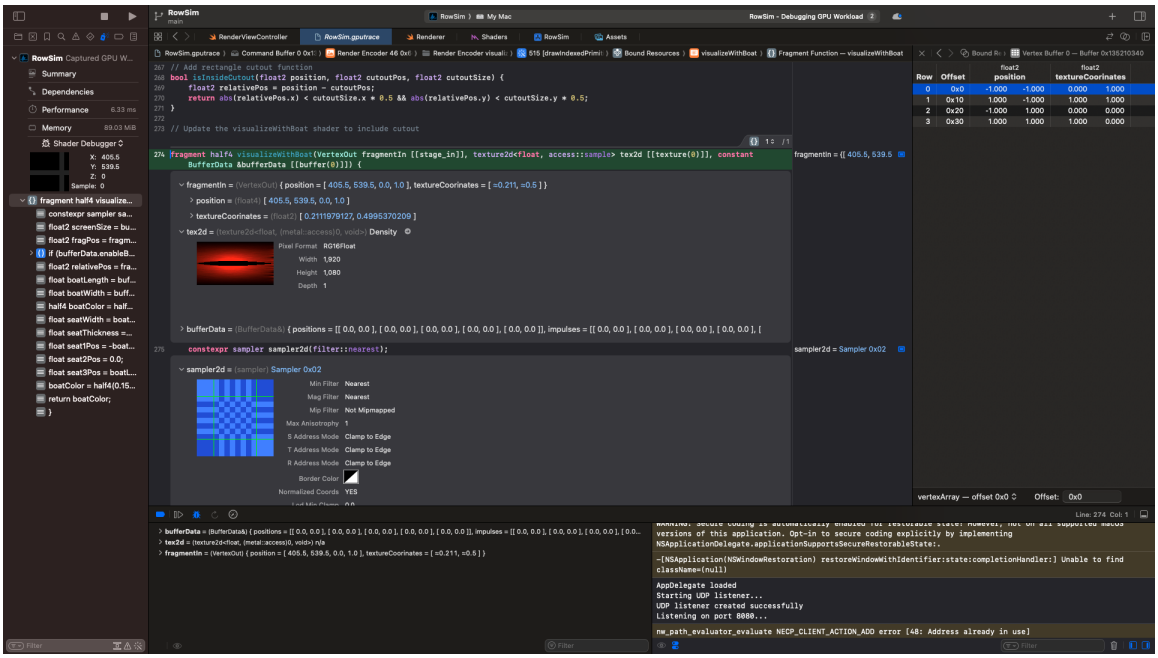


Figure A.1: A GPU frame capture taken during development, stopped inside `visualizeWithBoat` at fragment $(405.5, 539.5)$. The bound density texture reports pixel format `RG16Float` at $1{,}920 \times 1{,}080$, confirming the storage format and the once-allocated grid size described in this section. The bound-resource inspector at centre is where those values appear; the thumbnail beside them shows the density field itself, with the boat cutout as the dark lens and the wake shed from its ends. The captured workload runs at 6.33 ms with 89.03 MiB resident, both visible in the navigator at left.

A.3 Fluid and Wave Solver: Full Discretization

A.3.1 Advection

Velocity, density, and other transported fields are advanced with semi-Lagrangian advection, the unconditionally stable scheme introduced for real-time graphics by Stam (113, 114) and described in standard form by Bridson (22). For grid location $\mathbf{x}$, the shader traces backward

Table A.1: Notation used throughout this section. Fields are stored per texel on the simulation grid; subscripts i, j index that grid.

Symbol	Meaning
$\mathbf{u}$	Velocity field, in grid cells per frame
$\tilde{\mathbf{u}}$	Velocity after advection and forcing, before pressure projection
$\mathbf{u}_0$	Constant carrier flow representing the boat's motion through the water
q	Any advected scalar field (density, bloom energy, particle mass)
p	Pressure
α, β	Jacobi/Gauss–Seidel coefficients for the five-point Poisson stencil
ω	Scalar curl (vorticity), out-of-plane component
h	Wave height on the ripple field
c	Wave propagation speed, in grid cells per frame
d	Per-frame wave damping coefficient, $0 \leq d < 1$
s, s_r	Ripple-to-velocity coupling gains for the height gradient and for its per-frame change
Δt	Simulation timestep, fixed at one frame

through the velocity field plus a constant carrier flow $\mathbf{u}_0$ and bilinearly samples the source field at the traced position, excluding samples that fall inside obstacles to reduce density leakage through solids:

$$q^{t+1}(\mathbf{x}) = q^t\big(\mathbf{x} - \Delta t\,[\mathbf{u}(\mathbf{x}) + \mathbf{u}_0]\big). \tag{A.2}$$

Velocities are carried in grid cells per frame and Δt is fixed at one frame, so the timestep is unity in shader units and does not appear in the kernel; it is written here because the scheme is only unconditionally stable in the sense Stam intends when the backtrace is taken over an explicit interval. A centre-distance dissipation term and a low-speed edge-bleed term are applied afterward as deliberate, documented artistic stabilizers rather than physical effects.

A.3.2 Divergence and Pressure Projection

Divergence is assembled with centred first differences:

$$(\nabla \cdot \tilde{\mathbf{u}})_{i,j} = \tfrac{1}{2}\big[(u^*_{i+1,j} - u^*_{i-1,j}) + (v^*_{i,j+1} - v^*_{i,j-1})\big], \tag{A.3}$$

where the starred neighbour samples are the ordinary field values at solid-free cells, and are replaced by the local cell's own velocity component wherever the neighbour lies inside an obstacle. That substitution makes the differenced quantity vanish across a solid boundary, which is the discrete form of the no-flux condition and is what keeps the pressure solve

from driving flow into the hull. The substitution is four lines of the kernel, reproduced here because the phrase "starred neighbour" is doing real work and is easier to check than to describe:

DIVERGENCE_COMPUTE no-flux neighbour substitution — Shared/Shaders/FluidCompute.metal

```
float2 vc = velocity.read(gid).xy;
float vl = obstacles.read(left).y  > 0.5 ? vc.x : velocity.read(left).x;
float vr = obstacles.read(right).y > 0.5 ? vc.x : velocity.read(right).x;
float vb = obstacles.read(down).y  > 0.5 ? vc.y : velocity.read(down).y;
float vt = obstacles.read(up).y    > 0.5 ? vc.y : velocity.read(up).y;

float div = 0.5 * (vr - vl + vt - vb);
```

A neighbour inside a solid contributes the cell's *own* component, so the corresponding difference is identically zero. The pressure Poisson equation $\nabla^2 p = \nabla \cdot \tilde{\mathbf{u}}$ uses the standard five-point Laplacian; in the usual α, β parameterization of that stencil this is $\alpha = -(\Delta x)^2 = -1$ and $\beta = 4$, the four being the number of contributing neighbours on a two-dimensional grid. On Apple Silicon, `jacobi_rb_compute` performs an in-place Red–Black Gauss–Seidel sweep: one dispatch updates "red" checkerboard cells, the next updates "black" cells using the just-updated red neighbours, for 16 outer (red+black) iterations. On the fragment fallback path, pressure is solved with 20 ping-pong Jacobi iterations, since a fragment pass cannot express the same in-place checkerboard update directly; this ping-pong Jacobi formulation is the one established for GPU fluid solvers by Harris (52) and extended to three dimensions by Crane et al. (31). Both branches use a fixed iteration count rather than iterating to a convergence tolerance, and neither is a multigrid method (79); at RowSim's grid resolution a fixed sweep budget produced flow judged acceptable during development, and bounded per-frame cost was prioritized over exact convergence; the residual was not measured, so this is an engineering judgment rather than a characterized result. Velocity is then projected by subtracting the pressure gradient, $\mathbf{u}^{t+1} = \tilde{\mathbf{u}} - \nabla p$, with obstacle-aware non-penetration handling at user-painted solids.

A.3.3 Vorticity Confinement

Scalar curl is estimated from neighbouring velocity samples,

$$\omega_{i,j} \approx \tfrac{1}{2}\left[(v_{i+1,j} - v_{i-1,j}) - (u_{i,j+1} - u_{i,j-1})\right], \tag{A.4}$$

and a confinement force computed from the normalized gradient of $|\omega|$ is added back to velocity to counteract the numerical damping introduced by semi-Lagrangian advection (43, 116). Two refinements beyond the textbook formulation are included: obstacle-edge curl shedding, which converts tangential slip at the hull boundary into additional rotational forcing, and a partial-slip boat band, in which velocity inside the hull is forced toward a fraction of the carrier velocity rather than to zero, so flow appears to slide along the hull. Force magnitudes are capped per texel (approximately 7.2–7.8 near obstacles, 15.0 globally) to prevent solver instability from overlapping splats.

The slip band takes two different forms in the two kernels that write hull interiors, which is worth stating because a single quoted coefficient would misdescribe it. In the vorticity pass the scale is flat, $-\mathbf{u}_0 \times 0.62$. In the projection pass it is tapered by distance from the hull edge and boosted toward bow and stern, so the band runs from 0.46 at the edge to a capped 1.35 at the tips — that is, the extremities of the hull are driven *faster* than the carrier flow, which is what makes the ends visibly shed:

GRADIENT_COMPUTE hull interior, obstacle branch — Shared/Shaders/FluidCompute.metal

```
if (obstacles.read(gid).y > 0.5) {
    if (params.enableBoat && isInsideBoat(fragPos, ...)) {
        float edgeDist = -boatSignedEdge(fragPos, ...);
        float nX       =  boatLocalNormX(fragPos, ...);

        float taper    = smoothstep(0.32, 2.15, edgeDist);  // 0 at the edge, 1 deep inside
        float tipBoost = smoothstep(0.55, 1.00, abs(nX));   // 0 amidships, 1 at bow and stern
        float slipMix  = (0.46 + 0.62 * taper) * (1.0 + 0.90 * tipBoost);

        output.write(float4(-params.baseVelocity * min(slipMix, 1.35), 0, 0, 0), gid);
    } else {
        output.write(float4(0.0), gid);                     // user obstacle: no-slip
    }
    return;
}
```

Repeated geometry arguments are elided as `...` for legibility; nothing else is changed. The `else` branch is the ordinary no-slip condition, and the contrast between the two branches is the whole of the design decision described in Section 3.6: user-painted obstacles are walls, the boat is not.

A.3.4 Wave / Ripple Field

Ripples are a separate two-channel wave-height texture (current and previous height) evolving under an explicit damped wave equation,

$$h^{t+1} = (2-d)h^t - (1-d)h^{t-1} + c^2\nabla^2 h^t, \tag{A.5}$$

in which d damps the field toward rest each frame and c sets propagation speed, with a five-point Laplacian and a Courant–Friedrichs–Lewy (CFL) stability clamp $c \leftarrow \min(c, 0.707)$, since stability on a unit grid requires $c^2 < 0.5$. Obstacle interiors are zeroed (fixed boundary), top/bottom boundaries reflect, and left/right viewport edges are open (zero-height substitution) so waves can leave the domain. Ripple height gradients are coupled back into velocity as $\mathbf{v}_{\text{ripple}} = -s\nabla h + s_r\widehat{\nabla h}\,(h^t - h^{t-1})$, where the first term pushes fluid downhill along the wave surface and the second adds a contribution in the gradient direction proportional to how fast the height is changing, so a rising wavefront drives flow ahead of itself; capped at $6.0\times$ the configured ripple-fluid-strength on the compute branch ($3.0\times$ on the fragment fallback).

A.3.5 Bloom and Particles

A bloom-energy slab accumulates Gaussian splats at interaction and oar-trail positions, is advected with the velocity field, decays each frame (default 0.975, with a sub-perceptual floor below which values are zeroed), and is blurred through a half-resolution chain ($2\times$ downsample, then a 13-tap separable Gaussian). Particles, when enabled, are a separate slab advected leftward at the carrier velocity, seeded at the domain's trailing edge with hash noise, conveying flow speed as drifting foam-like structure. Both stages are compute-only and are omitted on the fragment fallback branch.

A.4 Compute / Fragment Branch Divergence

Table A.2 lists literal constants that differ between the two architecture branches. These differences are structural performance/quality trade-offs, not bugs, but they mean the two branches are not bit-identical. The branch is selected at compile time from the host architecture — compute on Apple Silicon, fragment on Intel — and is not itself written to the manifest; what the manifest records is the host and the build (`hostname`, `os_version`, `app_version`), from which the branch follows. For this study the question does not arise in practice: all seventeen sessions ran on the same Apple Silicon machine under the same

build, so every participant saw the compute branch and the fragment fallback was never exercised.

Table A.2: Compute (Apple Silicon) vs. fragment-fallback branch divergence.

Behaviour	Compute branch	Fragment fallback
Pressure solver	Red–Black Gauss–Seidel, 16 iterations	Jacobi, 20 iterations
Ripple injection radius	16.0	8.0
Ripple coupling multiplier	$\times 0.5$	$\times 0.3$
Ripple velocity cap	$6.0\times$ strength	$3.0\times$ strength
Bloom blur / particle advection	Present	Absent
Target frame rate	60 fps	30 fps

A.5 PM5 Telemetry: Service and Characteristic Layout

RowSim's native PM5 client connects to the Concept2 PM5 rowing service (`CE060030`) and subscribes to five characteristics, summarized in Table A.3. All payloads are little-endian: multi-byte values arrive least-significant byte first, which is what the parser assumes when it reassembles them.

Table A.3: PM5 BLE characteristics consumed by RowSim.

Characteristic	Key fields
General status (`CE060031`)	Elapsed time, distance, rowing state, stroke state, drag factor
Additional status 1 (`CE060032`)	Speed, stroke rate, heart rate, pace, average pace
Additional status 2 (`CE060033`)	Average watts, calories
Stroke data (`CE060035`)	Drive length/time, recovery time, stroke distance, peak/average drive force, stroke count
Additional stroke data (`CE060036`)	Instantaneous watts, stroke calories, stroke count

Drive events are deduplicated by a signature over (stroke state, drive time, recovery time, drive length, peak force); a session-level cooldown gate further suppresses duplicate drive-edge artefacts from asynchronous BLE characteristic arrival. PM5 connection is considered stale after 5.0 s without new data; the PM5 target-velocity contribution to the

physics controller (Section 3.9) is separately timed out after 2.0 s, so a brief telemetry stall does not immediately force a full disconnect classification.

A.6 Inertial Sensing: Firmware and Signal Layout

The handle-mounted inertial unit (Figure 3.9) runs separate firmware, maintained in its own repository (Section A.1). An ESP32 reads an ADXL345 accelerometer and an ITG3200 rate gyroscope over I^2C and fuses them with a complementary filter into a single orientation estimate, from which world-frame acceleration is derived and then integrated to velocity and displacement through a leaky second-order filter whose cutoff and Q are exposed as writable parameters. The device advertises itself on the network as a libmapper device named `ergometer`; RowSim's native libmapper provider subscribes to it by signal name, which is why a device can be swapped without touching the study-facing event model (Section 3.8).

Table A.4 lists the signals RowSim subscribes to. Each is written to `raw/events.jsonl` under its own `sourceStream` label, and all ten arrive together at roughly 15,600 records per session — an order of magnitude more traffic than the PM5 stream, which is why the raw log is the larger artefact in every bundle.

Table A.4: The inertial unit's libmapper signals as subscribed by RowSim. Vector length is the signal's dimensionality; orientation is a quaternion.

Signal	Len.	Unit	Quantity
`bodyFrame/acceleration`	3	m/s^2	Raw sensed acceleration
`bodyFrame/rotation`	3	rad/s	Raw angular rate
`worldFrame/orientation`	4	—	Fused orientation (quaternion)
`worldFrame/acceleration`	3	m/s^2	Acceleration, gravity removed
`worldFrame/rotation`	3	rad	Angular displacement
`worldFrame/velocity`	3	m/s	First integral, leaky
`worldFrame/displacement`	3	m	Second integral, leaky
`worldFrame/roll`	1	rad	Roll angle
`worldFrame/tilt`	1	rad	Tilt angle
`worldFrame/azimuth`	1	rad	Heading

Stroke detection from the inertial stream. A stroke edge is not read off a single threshold crossing. The unit's horizontal acceleration magnitude is tracked against a slowly

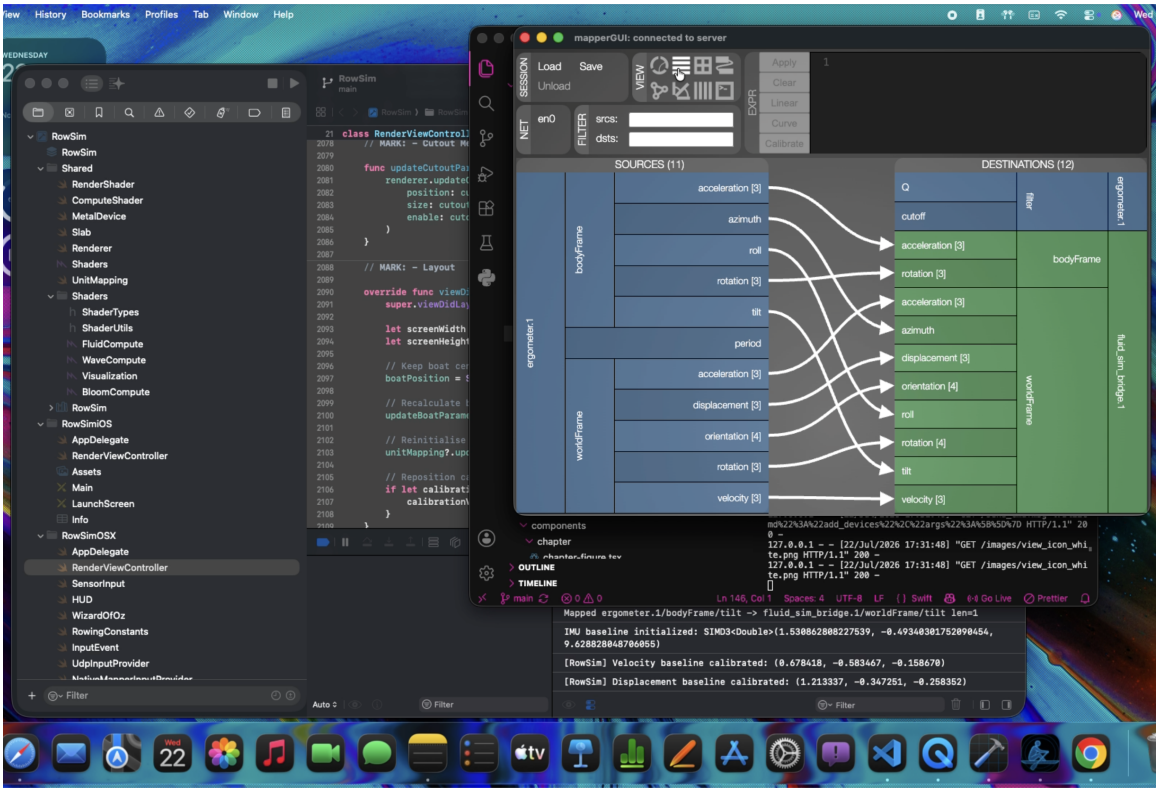


Figure A.2: The libmapper session graph, shown in `webmapper`. At left are the eleven signals published by the handle unit as device `ergometer.1`, grouped into `bodyFrame` and `worldFrame` with a separate `period` signal. At right are twelve destinations: the ten in green belong to `fluid_sim_bridge.1` and are the ones the table above lists, while `Q` and `cutoff` in blue are filter parameters on the ergometer device itself and are left unconnected. The curved connectors are the live maps. Signals are addressed by name, so replacing the sensing hardware means republishing the same names rather than changing the consumer. The console beneath the window shows the baselines being established at session start — inertial, velocity and displacement in turn — which is the calibration described in Section A.6.

adapting baseline, and engagement uses hysteresis: motion must exceed an engage threshold, or a jerk threshold, on several consecutive samples before the detector arms, and must fall below a lower disengage threshold before it can re-arm. Two refractory intervals then apply, both on this path: a source-specific 0.6 s and a further 0.4 s guard named `unifiedMinStrokeInterval` in the source. The name is misleading and is worth correcting here, since a reader will meet it: the constant is tested only on the inertial branch, and the PM5 and velocity branches carry their own guards of 0.25 s and 0.3 s/0.25 s respectively (Section 3.9). Because PM5 owned triggering in every session of this study, neither of the intervals in this table was ever the operative one. Table A.5 lists the values, all of which are recorded per session in `manifest.json` and were identical across all seventeen.

Table A.5: Inertial stroke-detection and calibration parameters, identical in all seventeen sessions. The engage and disengage thresholds differ deliberately: a single threshold would re-trigger on the noise around it, so the detector arms high and releases low.

Parameter	Value	Role
Engage threshold	0.8	Arms the detector
Disengage threshold	0.3	Releases it; the gap is the hysteresis
Jerk threshold	1.5	Alternative arming path on sharp onset
Deadband	0.2	Motion below this is ignored
Still threshold	0.2	Defines "at rest" for baseline adaptation
Jerk-still threshold	0.5	As above, on the jerk channel
Baseline EMA α	0.02	Rate at which the baseline tracks drift
Calibration samples	15	Still samples required to set the baseline
Source refractory	0.6 s	Minimum interval between inertial strokes
Unified refractory	0.4 s	Minimum interval between admitted triggers on this path only; PM5 admission uses a 0.25 s cooldown

The baseline is set from the first still samples and then adapts continuously while the unit is at rest, which is what makes the detector tolerant of the sensor drift that accumulates over a ten-minute session. It also means the detector is stateful: an inertial stroke series is not reproducible from the raw acceleration alone without replaying the same baseline history, which is one reason the analytical stroke series in Chapter 5 is derived from the ergometer rather than from this stream.

A.7 Runtime Configuration Surface

A small number of environment variables change which input path and network interface RowSim uses at launch, and therefore change what a session actually exercises. Table A.6 lists them together with their defaults. Their values are captured in each session's `manifest.json`, so the configuration a given session ran under can be reconstructed after the fact rather than assumed.

Table A.6: Environment variables affecting session behaviour, recorded per session.

Variable	Default	Effect
`ROWSIM_INPUT_UDP`	unset	`1` selects the UDP input provider instead of native libmapper
`LIBMAPPER_INTERFACE`	`en0`	Network interface used for libmapper discovery
`ROWSIM_MAPPER_DEBUG`	unset	Enables verbose native-mapper lifecycle logging
`ROWSIM_SESSIONS_DIR`	platform default	Overrides where session bundles are written

A.8 Additional Installation and Output Views

Figure 3.13 shows nine visual configurations at a size sufficient to tell them apart but not to see what each one does to the flow. Two are worth reproducing larger, because in each case what distinguishes the configuration is not its colour: the caustic layer's response to handle rotation (Figure A.3), and the blue-on-black configuration seen from the four viewpoints a rower and an observer actually occupy (Figure A.4).

A.8.1 The caustic layer, and what participants made of it

The caustic layer is worth separate comment because it is the one visual element driven entirely by the inertial unit rather than by the ergometer, and because what participants made of it is not what the design anticipated. It has no enable flag: the final pass applies it to every fragment outside the hull, unconditionally. What sets its strength is the *rate of change* of handle azimuth, accumulating while the handle turns and decaying back toward zero once it stops. Small rotations produce a low, diffuse brightening; sharper ones produce a brighter and more localised response.

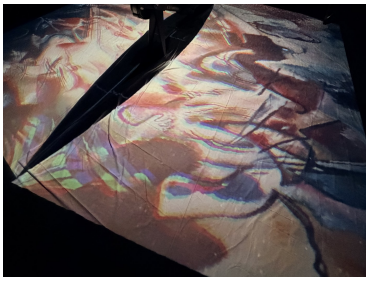
It was driven continuously. Azimuth — one of the eleven signals mapped in Figure A.2 — was sampled a median of 40,693 times per session and varied in every one of the seventeen, with a within-session standard deviation of 19.6 units at the median and a 95th-percentile rate of change of 42 units per second. A rowing stroke is not a rigid translation: the handle rolls a little through the drive and rather more at the catch and the release, which is enough to keep the layer live throughout.

Participants noticed that tilting the handle did something, and none of them had a word for what. The clearest description of the caustic layer itself comes from P15, who read it as a change of colour rather than of light — "when I moved the pedals a little bit to the right or

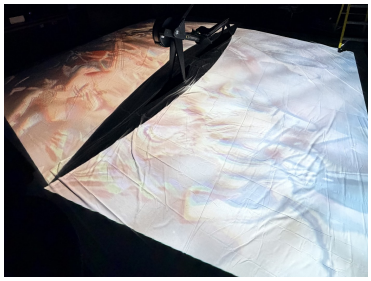
left, it changed the colors" — which is what a brightening toward white looks like when the field beneath it is magenta. P8 went looking for a tilt response deliberately, without finding a name for what she found: "Can I vary, if I sort of tilt the handle to make one side have a stronger stroke than the other?"

P7's account is the most interesting of the three, and it belongs to a different layer than the other two. He described tilting the handle and seeing "like droplets on the water," which he judged "very different and creative of the RowSim, so you know when you're out of balance and stuff like that." Droplets are the vocabulary of the ripple heightfield rather than of the caustics, and the ripple layer is in fact roll-driven: a roll change past threshold spawns a Gaussian splash to the corresponding side of the hull, rate-limited so a steady stroke does not saturate it. What P7 reported is therefore an accurate reading of a real coupling. What makes it worth reporting is the referent he reached for. Balance is not a quantity RowSim computes or intends to display, and on the water it is exactly what handle roll indicates — so an unlabelled response to handle rotation was read as the thing handle rotation actually means to a rower.

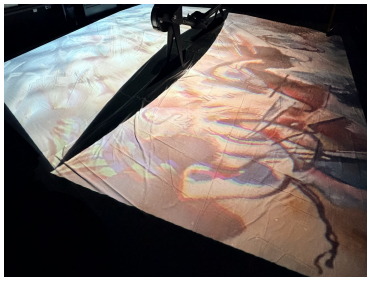
Two things follow. The first is that a display element can be legible without being nameable: three participants perceived a rotation-driven response and none used the vocabulary the implementation does, yet P7 arrived at the referent the coupling was physically about. The second is a caution about the sharpness of the effect. Its amplitude scales with how abruptly the handle turns, and a rower holding a consistent technique produces a steady stream of small rotations rather than occasional large ones, so for most of a session the layer contributes an even wash rather than a distinct event. That, together with the contrast budget discussed in Section 6.3, is the likely reason it registered with three participants rather than seventeen. The caustic parameters — bias strength and falloff — are the only display settings in this appendix the manifest does not record, so their exact values during the study cannot be recovered.



(a) Caustic layer off

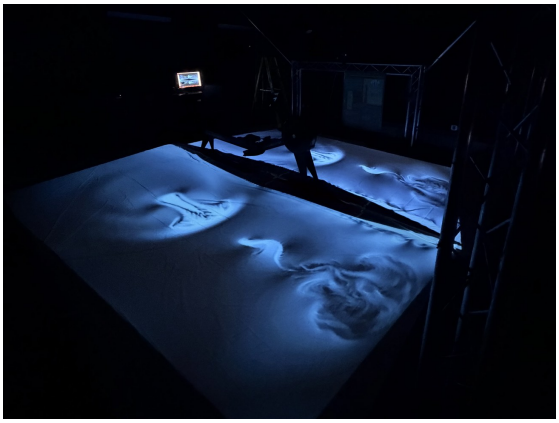


(b) Tilted one way

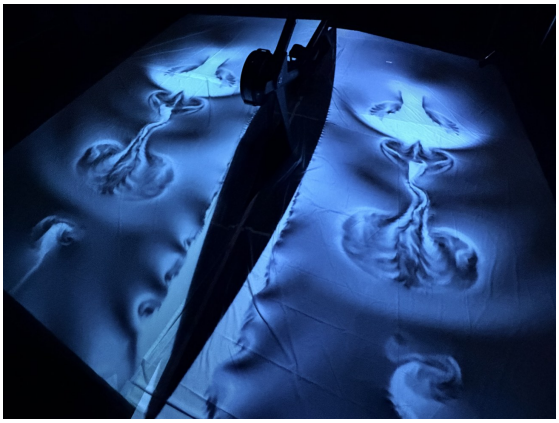


(c) Tilted the other

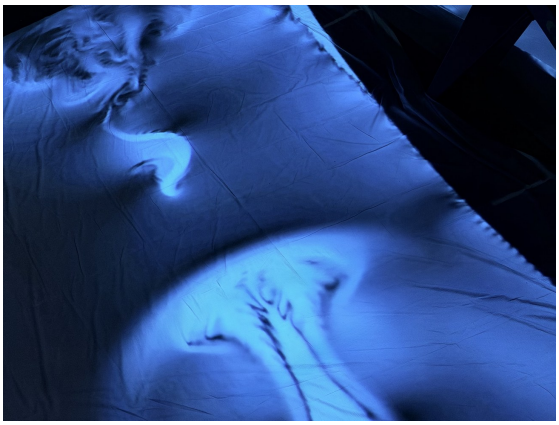
Figure A.3: The caustic layer responds to handle rotation rather than to colour: (a) with the layer suppressed, and (b)–(c) with the handle rotated in opposite directions, the bright region moves across the field while the palette stays put. The *image background* shown here was not used in the study — `background_image_enabled` is false in all seventeen manifests — but the caustic response itself was active throughout, over the plain white background instead. Section A.8.1 reports what participants made of the effect, which was not what the design anticipated.



(a) Full projected area



(b) Oblique



(c) Wake structure

(d) From the seat

Figure A.4: Blue-on-black mode from four viewpoints: (a) full projected area, (b) oblique view, (c) wake detail, and (d) seated view. Only ground and tint differ from the study configuration; this mode was not used in participant sessions.

A.9 Session Logging Artefacts

One property of the manifest matters when reading it. It is written when the session is closed, alongside the SHA-256 checksums, which can only be computed once the artefacts are complete. The `simulation_state` block it carries is therefore a snapshot of the configuration at that moment. Most of those fields cannot change during a session, but a few are adjustable at runtime from the keyboard, and for those the manifest records the state at write time rather than a guarantee of the state throughout the rowing segment.

Each session produces a self-contained bundle: `manifest.json` (session identity, runtime configuration, software snapshot, SHA-256 artefact checksums), `raw/events.jsonl` (an append-only stream of every sensor, PM5, and configuration-snapshot record with source and ingest timestamps), and `formatted/sensor_data.csv` (the same data bucketed at 50 Hz into a wide-format table for analysis). This structure is what makes the branch- and configuration-recording described above auditable after the fact, rather than only asserted (Figure A.5).

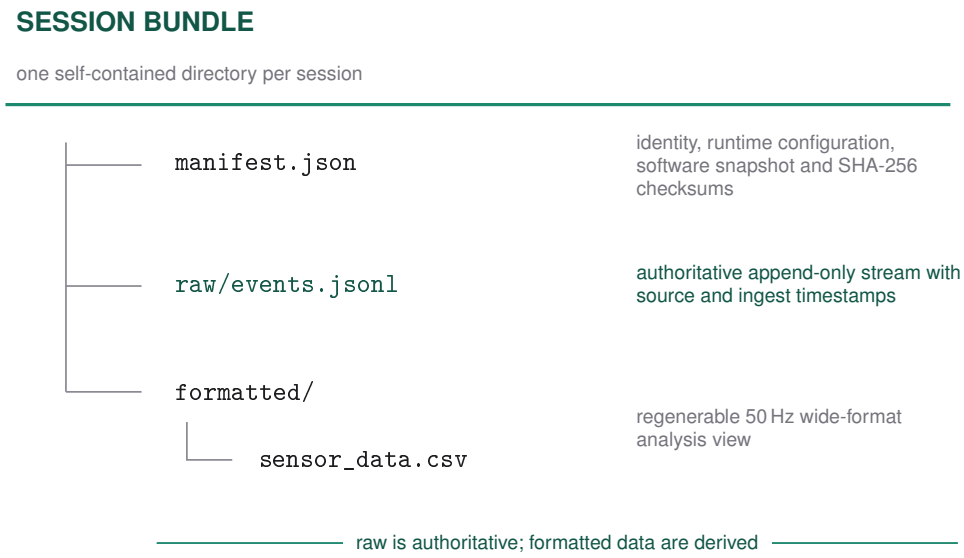


Figure A.5: Self-contained session logging bundle. The append-only raw stream is authoritative, the analysis CSV is regenerable, and manifest checksums expose subsequent alteration.

Appendix B

Instruments and Scoring

This appendix reproduces the exact items administered and the exact scoring applied, so that the values reported in Chapter 5 can be checked without recourse to the ethics submission. All Likert items used a five-point agreement scale (1 = strongly disagree, 5 = strongly agree).

B.1 Engagement Questionnaire

The twelve items of the User Engagement Scale — Short Form (89), worded for the rowing context. Table B.1 lists each item exactly as administered beside its counterpart in the published scale, so the correspondence can be checked directly.

Scoring. Items 3 and 6 are reverse-scored. Each subscale is the mean of its constituent items; the overall score is the mean of all twelve items after reversal. This follows the published short form's scoring procedure, which directs that PU-S.1–PU-S.3 be reverse coded, that each subscale be the sum of its items divided by three, and that the overall score be the sum of all items divided by twelve (89). The divisors differ here because the administered instrument is not distributed 3/3/3/3: focused attention carries two items and perceived usability four, so each subscale is divided by its own item count rather than by three. The overall divisor of twelve is unchanged, since twelve items were administered.

Relationship to the published short form. Nine of the twelve published items were administered, with wording set to the rowing context — a change the administration instructions explicitly allow, giving the example of rewording "I felt frustrated while using this Application X" to name the system under study. Three were not administered, and each was replaced by an item measuring something else. FA-S.2 ("The time I spent… just slipped away") and PU-S.3 ("Using this Application X was taxing") gave way to two positively worded items on perceived control and confidence, assigned here to perceived usability on the grounds that both concern the participant's sense of competent operation. RW-S.3 ("I felt interested in this experience") gave way to "The experience felt valuable to me": interest and value are related but not the same construct, so this substitution should be read as changing what the reward subscale measures rather than as rewording it. The four dimensions are

Table B.1: Each item as administered, with its published UES-SF counterpart, the subscale used in analysis, and the scoring direction. Reverse-scored items are transformed as $6 - x$ so that higher always indicates greater engagement. The published scale's administration instructions permit wording changes for context; items 4, 5 and 12 have no published counterpart and stand in place of FA-S.2, PU-S.3 and RW-S.3.

#	Subscale	Item as administered	UES-SF	Dir.
1	Focused attn.	I lost myself in this experience.	FA-S.1	+
2	Focused attn.	I was absorbed in the activity.	FA-S.3	+
3	Perceived usab.	I felt frustrated while using the system.	PU-S.1	−
4	Perceived usab.	I felt in control of the interaction.	—	+
5	Perceived usab.	I felt confident using the system.	—	+
6	Perceived usab.	I found the system confusing.	PU-S.2	−
7	Aesthetic appeal	I found the interface visually appealing.	AE-S.3	+
8	Aesthetic appeal	The interface was attractive.	AE-S.1	+
9	Aesthetic appeal	The interface was aesthetically pleasing.	AE-S.2	+
10	Reward	Using the system was worthwhile.	RW-S.1	+
11	Reward	My experience was rewarding.	RW-S.2	+
12	Reward	The experience felt valuable to me.	—	+

populated 2/4/3/3 rather than 3/3/3/3, and one of the three reward items is not a published one.

Two consequences follow for interpretation. Subscale and overall means should be read against this instrument rather than compared numerically against published UES-SF results, which is why Chapter 5 reports item-level means alongside them. And the focused-attention dimension is the most affected, since it rests on two items rather than three — which matters because Section 5.2 shows participants interpreted those two items differently from one another. Neither consequence affects the internal validity of the comparisons reported, all of which use the same instrument throughout.

B.2 RowSim Experience Questionnaire (REQ)

A custom instrument developed for this study. Twenty Likert items followed by four free-text prompts. Items marked − are reverse-scored in the composite. Item 11, on physical effort, is excluded from the composite entirely: high physical effort is expected in a rowing task

and is neither straightforwardly positive nor negative. The responses suggest the item was not read as being about rowing at all. It returned a mean of 1.88 — broad disagreement that the experience required much physical effort — from participants who had just rowed for ten minutes, and the elite group, which produced 3.7 times the power of the casual group, disagreed more strongly than the casual group did (1.40 against 2.08). Read as a question about exertion those numbers are not credible; read as a question about the interface they are unremarkable. Asking whether a physically demanding activity was physically demanding turns out not to discriminate, and the phrase "the experience" invited participants to answer about the display instead. The item is retained in the reported table for completeness and carries no analytic weight.

The twenty Likert items are reproduced in Table B.2 as administered. They were written in four thematic blocks of five — comprehension and legibility (1–5), engagement and attention (6–10), effort and interference (11–15), and preference and perceived value (16–20) — and were administered in that order. The blocks are a drafting device rather than validated subscales: no factor structure was hypothesized or tested, and no block is scored separately anywhere in this thesis. They are stated here because the ordering is not random and a reader reconstructing the instrument should know that related items were adjacent, which can induce local consistency in responses.

Free-text prompts. (i) What did you like most about the experience? (ii) What did you like least, or what frustrated you during the experience? (iii) How could the projected visuals and/or interaction be improved? (iv) Any further comments?

Scoring. The composite is the mean of the nineteen scored items after reversal of those marked −. Item 11 is excluded. Item-level means are reported in Section 5.2 alongside the composite, and the item-level values are the primary reporting level for this instrument.

Caveat. The REQ is a study-specific instrument with no independent validation. Its composite is used as a descriptive summary only, and no inference in this thesis rests on treating it as a validated scale.

Table B.2: The twenty REQ Likert items as administered, with scoring direction. "Excl." marks the item omitted from the composite.

#	Item	Dir.
1	I found the experience easy to understand.	+
2	I found the experience confusing.	−
3	The projected visuals matched my actions in a meaningful way.	+
4	I was able to quickly learn how to affect the projected visuals.	+
5	At times, I felt unsure about what the projected visuals meant.	−
6	The experience made me want to row more.	+
7	I felt distracted by the projected visuals.	−
8	I felt absorbed in the experience while rowing.	+
9	The experience kept my attention throughout.	+
10	I found the experience boring or disengaging.	−
11	I felt that the experience required a lot of physical effort.	Excl.
12	I had difficulty maintaining my rowing rhythm because of the projected visuals.	−
13	I felt that the experience required a lot of mental effort (e.g. interpreting its feedback).	−
14	The projected visuals helped reduce the monotony of rowing on an ergometer.	+
15	The projected visuals made rowing feel more tiring than usual.	−
16	Overall, I enjoyed the experience.	+
17	I would prefer using the projected visuals rather than a standard rowing machine display.	+
18	I would choose to use a standard rowing machine instead if given a choice.	−
19	I would recommend this experience to others.	+
20	I believe this experience could be useful for training or rehabilitation.	+

B.3 Internal Consistency of the Two Scales

Because the engagement instrument was adapted (Section B.1) and the REQ was written for this study, neither carries published reliability as administered. Cronbach's α was therefore computed on the seventeen complete responses for each scale and each subscale, after reversal. Table B.3 reports the results with percentile bootstrap intervals over 5,000 resamples of participants.

Three of these are unremarkable. Aesthetic appeal and reward are internally consistent at the level published scales typically report, the overall engagement score is acceptable,

Table B.3: Cronbach's α for each scale and subscale as administered, $n = 17$, with 95% percentile bootstrap intervals. The intervals are wide because seventeen respondents is a small basis for a reliability estimate; they are reported so that the width is visible rather than implied.

Scale	Items	α	95% CI
Engagement scale, overall	12	.84	[.64, .92]
Focused attention	2	.04	[−.93, .63]
Perceived usability	4	.64	[−.24, .79]
Aesthetic appeal	3	.87	[.58, .95]
Reward	3	.90	[.77, .97]
REQ composite	19	.82	[.55, .90]

and the REQ composite holds together well enough to be used as the descriptive summary it is used as. Perceived usability at $\alpha = .64$ is weaker, which is what the substitution described in Section B.1 would predict: the subscale mixes two published items on frustration and confusion with two substituted items on control and confidence, and those are related but not identical concerns.

The focused-attention result is not a minor irregularity. At $\alpha = .04$, with an inter-item correlation of $r = .02$, the two items assigned to that dimension share essentially none of their variance. Participants who reported losing themselves in the experience were no more likely than anyone else to report being absorbed in the activity. A two-item subscale cannot absorb this: with α at that level the subscale mean is not measuring one thing, and the focused-attention figure reported in Table 5.1 should be read as the average of two separate observations rather than as a scale score.

This is a limitation of the instrument as administered, and it is stated here as one. But it is also the sharpest available evidence for the argument in Section 6.4, which was built on the standard deviation of the "lost myself" item and on participants asking aloud what the word was meant to mean. The reliability coefficient adds something those cannot: it establishes that the two items were not merely answered with different spreads but were answered *independently of each other*. The item standard deviations make the same point from the other direction. Across all twelve items the spread runs from 0.24 to 0.71 — with one exception. "I lost myself in this experience" sits at 1.44, double the next widest item in the instrument, while its partner "I was absorbed in the activity" sits at 0.62, squarely mid-pack. The same seventeen people, asked about the same ten minutes, agreed closely on one of these and split almost evenly on the other.

Two cautions on how far this can be taken. A two-item α is a rescaled correlation and inherits all of that statistic's instability at $n = 17$; the interval runs from $-.93$ to .63 and

is consistent with a moderate positive association as well as with none. And low internal consistency has more than one possible cause: the two items may be tapping genuinely distinct constructs, or "lost" may be ambiguous in the way Section 6.4 argues, or both. The reliability figure cannot distinguish these. What it does establish is that the two items did not behave as one measure in this sample, which is a precondition for the argument rather than a proof of it.

B.4 Interview Guide

The post-session interview was semi-structured: the prompts below were asked of every participant, with follow-up questions improvised in response. Duration therefore varied considerably between participants.

1. What stood out to you most about the RowSim feedback?
2. What aspects of the visuals did you find helpful during rowing?
3. Did the visuals motivate you, engage you, or affect your focus in any way? How?
4. Were there any moments when the visuals confused you or were difficult to interpret?
5. Did the visuals affect your rowing rhythm or technique in any way — positively or negatively?
6. At any point, did the visuals feel distracting or overwhelming?
7. What did you like least about the system, or what frustrated you?
8. If you could change or improve one thing about the visuals, what would it be?
9. Is there anything else you would like us to know?

Prompts 4, 6 and 7 explicitly invite criticism. This was deliberate, given that the interviewer was also the system's designer (Section 4.8), and the responses to those prompts supply most of the material in Themes 4 and 5.

B.5 Demographic and Background Questionnaire

Administered before the session, on paper, and used for the sample description in Section 4.2 and for the group assignment described in Section 4.2. Table B.4 lists each field and its response format.

Table B.4: Fields collected before the session, with response format.

Field	Response format
Age	Band
Gender	Free text
Occupation	Free text
Rowing background	Casual *or* elite/competitive
Years of rowing experience	Band
Typical weekly rowing frequency	Band
Usual rowing settings	Multiple choice: indoor erg, on-water, rowing club, gym, home, other
Prior use of technology-based rowing feedback	Yes/no, with free text naming the system
Comfort with new digital systems or interactive displays	Five-point scale
Prior participation in sports, HCI or exercise-technology research	Yes/no
General weekly physical activity level	Band
Any condition relevant to participant comfort	Optional free text

The rowing-background item is the one that carries analytic weight: it is the basis of the group assignment, and it was checked against the experience and frequency items rather than taken alone, on the criteria stated in Section 4.2.

B.6 Research Ethics Board Approval

The study was reviewed under delegated review by the Dalhousie University Social Sciences and Humanities Research Ethics Board and approved on 10 February 2026 under file 2025–8251, valid for twelve months (Figure B.1). All seventeen sessions were run between June and August 2026 and therefore fall inside the approval period. The project title on the approval, *RowSim: A Mixed Reality Rowing System with Ambient Visual Feedback*, predates the present title; the protocol is the one reported in Chapter 4.

DALHOUSIE UNIVERSITY

Social Sciences and Humanities Research Ethics Board
Letter of Approval

February 10, 2026

Arsh Shah
Computer Science\Computer Science

Dear Arsh,

REB #: 2025-8251
Project Title: RowSim: A Mixed Reality Rowing System with Ambient Visual Feedback

Review Type: Delegated Review
Effective Date: February 10, 2026
Expiry Date: February 10, 2027

The Social Sciences and Humanities Research Ethics Board has reviewed your application for research involving humans and found the proposed research to be ethically acceptable in accordance with the *Tri-Council Policy Statement Ethical Conduct for Research Involving Humans*. This approval will be in effect for 12 months as indicated above. This approval is subject to the conditions listed below which constitute your on-going responsibilities with respect to the ethical conduct of this research.

Sincerely,

Dr. John Cameron
Chair, Social Sciences and Humanities Research Ethics Board
Dalhousie University

Funding NSERC RGPIN-2021-03845

Figure B.1: Letter of approval, Dalhousie University Social Sciences and Humanities Research Ethics Board, 10 February 2026.