Innovating inside a music-distribution company

An insider action-research study of how TDMusic decided whether, what and how to build an adaptive-audio wellness product (Lilt)

Applied management dissertation

Executive MBA · emlyon business school

Author: **Qiao Lu** · Supervisor: **[to confirm with the author]**

Draft of 6 September 2026 · Submission due before November 2026

All product language in this dissertation is general-wellness language. No diagnostic or therapeutic claim is made anywhere in this document or in the product it describes.

Acknowledgements

[to be written by the author.]

Abstract

An insider action-research study of how a mid-sized, profitable music-distribution company decided whether, what and how to build an adaptive-audio wellness product. TDMusic Inc. owns a validated engine — demand signal, AI-assisted production, measured library — and the question is whether it transfers to a health-adjacent need. Three questions organise the study: whether a sizeable, reachable, under-served need exists; whether the firm can serve it with its assets, measurably and with a viable business model; and how a firm of this type should organise and decide such an exploration. The method pairs design thinking, which generates and shapes the options, with a pre-registered Bayesian belief register that decides among them — priors with written rationales, likelihood ratios from a published rubric, GRADE-style shrinkage, and thresholds fixed before the data. The need is real, frequent and under-served in one specific way: stopping racing thoughts is the only outcome clearing the opportunity threshold. But it is state-dependent, and so is the content that serves it. In the acute moment listeners prefer engineered neutral sound to real music; the ordering reverses for wind-down. That runs against the firm's principal asset. The pre-registered content pivot fired, and the product became two modes rather than one. Three of five ability conditions clear their gates; the two efficacy conditions cannot be met without a pilot not yet run, so the verdict is NOT YET. The contribution is an auditable decision apparatus, an operationalisation of discovery-driven planning, demonstrated where the evidence turned against the firm's own resource.

Word count: 250.

Keywords: corporate entrepreneurship; ambidexterity; design thinking; discovery-driven planning; Bayesian decision-making; pre-registration; adaptive audio; digital wellness; insider action research.

Executive summary

The question. TDMusic Inc. is a profitable digital-content distribution company — reported revenue $10.5 M, net income $2.2 M, 52 employees — whose engine turns a demand signal into an AI-produced, measured, distributed library. That engine has been proven once outside music, in a Kindle Direct Publishing pilot. This dissertation asks whether it transfers to anxiety-relief audio, and reports the exploration that answered the question.

The method. A three-month, $10–20k ring-fenced design-thinking phase, run as insider action research by the firm's founder, with a decision layer written down before the data: eight beliefs, each with a prior and a written rationale; every piece of evidence converted into a likelihood ratio through a published rubric and shrunk by a published evidence-quality rule; and four gate rules — proceed, content pivot, form pivot, kill — expressed as posterior thresholds fixed in advance. The two efficacy thresholds were deliberately placed above what the meta-analytic literature alone can reach, so that a PROCEED decision cannot be cleared by literature review.

What the evidence said. The need is real and frequent — 87 per cent report needing a calming moment at least weekly, 90 per cent already use audio to get it — and under-served in one specific way: stopping racing thoughts is the only outcome above the opportunity threshold, and 38 per cent of wearable owners do nothing at all after a high stress reading. But the need is state-dependent. In the acute scenario 24 per cent chose a real song against 46 per cent choosing neutral sound; for wind-down the ordering reverses to 47 against 19; the within-person shift is significant, and the effect is sharper in the more anxious segment. That result runs directly against the firm's principal asset, its catalogue and the rights to it.

What was decided. The content-pivot rule fired mechanically: the acute product becomes engineered neutral sound and real music becomes the wind-down mode, so the product is two modes rather than one, and the acute mode requires no catalogue rights at all. Three of the five ability conditions clear their gates — engagement, willingness to pay, and a rights position that is more likely than not pending a legal check. The two efficacy conditions do not, and cannot, without the pilot. **The recommendation is NOT YET.**

What it cost and what it bought. The concept the firm's own assets would have selected — calm, AI-adapted versions of its own catalogue, sold as a consumer subscription — is the concept the completed register ranks last, at 26 per cent, because it depends on five beliefs at once. The lowest-cost concept, an artist calm-content line through the firm's existing distribution, ranks highest at 65 per cent. What the firm acquired is a reusable apparatus rather than a product: the next vertical costs it the fieldwork, not the machinery.

What remains. Run the legal check first — one to two weeks, almost no cost, and the only pending evidence graded decisive. Rewrite the landing-page copy, which still promises the pre-pivot proposition, before reading any conversion figure from it. Then run the efficacy pilot, apply the observed likelihood ratios, and write the decision memo. A documented KILL remains a valid outcome, and the ring-fenced budget exists to make that outcome survivable.

The largest caveat, stated at the front. The survey and focus-group figures throughout are the output of the project's pre-registered analysis pipeline as published on the project site, not field-work, and the efficacy pilot had not been run at the time of writing. Both are flagged at every point of use.

Table of contents

Printed page numbers appear at the foot of every page, and a running short title at the head. This contents list carries no page references: the build has no cross-reference pass, so section-to-page mapping is done by the reader from the running numbers. See `ASSEMBLY_REPORT.md`.

List of figures

List of tables

Abbreviations

ANCOVA	analysis of covariance
ARPU	average revenue per user
ARR	annual recurring revenue
B2B / B2B2C / B2C	business-to-business / business-to-business-to-consumer / business-to-consumer
BMC	Business Model Canvas
CAC	customer acquisition cost
CAGR	compound annual growth rate
CI	confidence interval
DiGA	*Digitale Gesundheitsanwendung* — a reimbursed German digital health application
DSP	digital service provider
DT	design thinking
DTx	digital therapeutic
E0 / E1 / E2 / E3 / E4	the calibration study, the efficacy pilot, and the three staged learning-loop experiments
FDA	United States Food and Drug Administration
FTC	United States Federal Trade Commission
GAD-2	Generalised Anxiety Disorder 2-item scale
GDPR	General Data Protection Regulation
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HR	heart rate
HRV	heart-rate variability
ICC	intraclass correlation coefficient
IPP	indifference price point
JTBD	jobs to be done
KDP	Kindle Direct Publishing
LR	likelihood ratio
LTV	lifetime value
LUFS	loudness units relative to full scale
MAPE	mean absolute percentage error
MVP	minimum viable product
NICE	National Institute for Health and Care Excellence (United Kingdom)
NMPA	National Medical Products Administration (China)
ODI	outcome-driven innovation
OPP	optimal price point
PEOU	perceived ease of use
PLS-SEM	partial least squares structural equation modelling
PMC	point of marginal cheapness
PME	point of marginal expensiveness
PSM	price sensitivity meter
PSS-4	Perceived Stress Scale, four-item form
PU	perceived usefulness
PWA	progressive web application

RA	research assistant
RCT	randomised controlled trial
RMSSD	root mean square of successive differences (a heart-rate-variability measure)
ROI	return on investment
RQ	research question
SaMD	software as a medical device
SDK	software development kit
SDNN	standard deviation of normal-to-normal intervals (a heart-rate-variability measure)
STAI-S / STAI-6	State–Trait Anxiety Inventory, state scale / its six-item short form
SWOT	strengths, weaknesses, opportunities, threats
TAM	total addressable market (chapter 3) *and* technology acceptance model (chapters 4–5); the two senses are distinguished by context at every use
SAM / SOM	serviceable addressable market / serviceable obtainable market
TME	Tencent Music Entertainment
VPC	Value Proposition Canvas
VRIO	valuable, rare, inimitable, organised (resource-based analysis)
WHO	World Health Organization
WTP	willingness to pay

1. Introduction

> **What this chapter establishes.** That the question before TDMusic is a corporate-innovation question inside a going concern rather than a start-up pitch; that the firm already owns a validated engine for turning a demand signal into a measured library of content; that pointing that engine at a health-adjacent need raises three researchable questions — need, ability and process; and that the whole exploration ran inside a declared boundary of general-wellness positioning, an overseas-first market, $10–20k and no more than three months.

1.1 Context: TDMusic and a proven engine

TDMusic Inc. (太簇音乐, Tempered Digital Music) is a technology-driven digital-content distribution and monetisation company headquartered in Beijing and operating across the United States, Europe, Japan and China. Company documents report revenue of $10.5 M, net income of $2.2 M, a $70 M valuation and 52 employees; an angel round led by Challenger Capital; and a market position described as top-three music distributor in China, Tier-1 YouTube partner and top supplier to Tencent Music. Catalogue figures dated 2022 record approximately 106,000 songs in reserve and 75,000 released, of which roughly 85 per cent are owned or directly licensed; 1.05 billion streams between January and July 2022; and distribution to more than 200 countries through 220+ digital service providers (TDMusic Inc., 2022–2026).

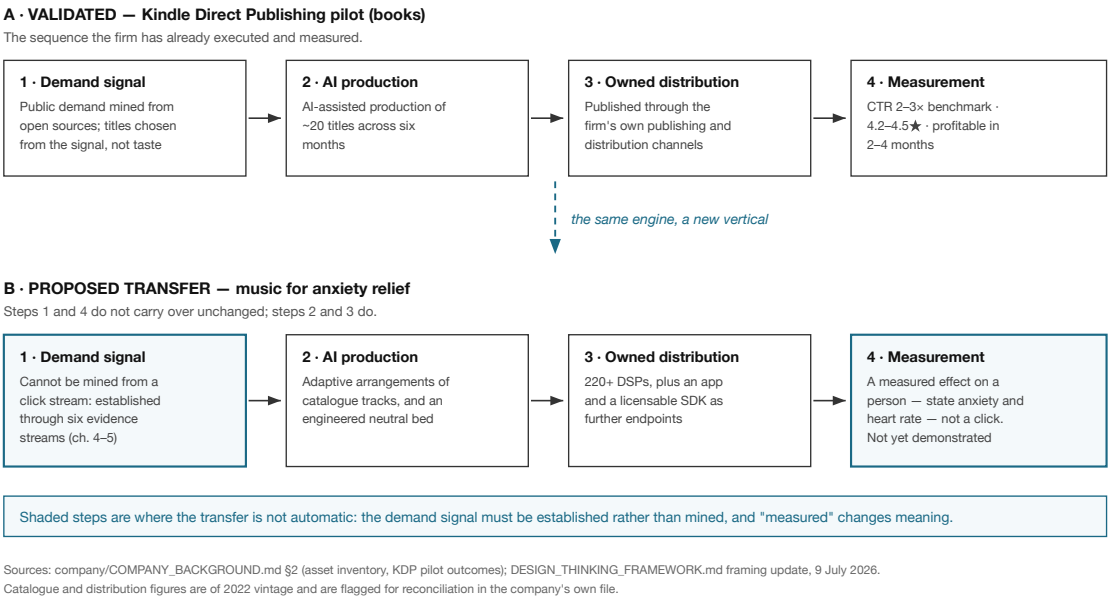
Four of those figures are flagged for reconciliation in the company's own file: a single founding date (2020 for the Chinese entity against March 2021 in the business-model deck), a single region count (four regions against five countries), the catalogue and stream counts, which are three years old, and the scope of the Kurt Hugo Schneider artist relationship (China agency operations against distribution rights). This dissertation reports each company figure with its vintage, uses no company figure inside an arithmetic result, and treats none of them as current. Closing that checklist is a condition of submission, and the current value of each of the four flagged items is **[to confirm with the author]**.

What matters for this study is less the size of the firm than the shape of its capability. Four assets are documented. First, a published recommendation and distribution algorithm — a scenario- and preference-based auto-matching engine described in a peer-reviewed article in *PeerJ Computer Science* (2021), whose byline the company file records incompletely, so the full author list is **[to confirm with the author]** — which the firm credits with an order-of-magnitude marketing-return advantage and which underlies a claimed per-song revenue roughly fourteen times the average of larger competitors in a 2022 internal comparison. Second, an AI-assisted production pipeline covering covers, audio artwork, intro content, voice synthesis and video, which the firm's own business-model deck singles out as strongest in the meditation and lo-fi genres. Third, rights and catalogue infrastructure: a majority-owned or directly licensed library with royalty administration, and composition rights covering adaptation for part of the catalogue. Fourth, distribution and platform relationships that reach 220+ digital service providers and endpoints as varied as Peloton, Tesla

and Virgin Airlines — a day-one global launch capability, and a set of wellness-adjacent business-to-business-to-consumer precedents already inside the house.

Those four assets have already been combined once, in a different medium, and the combination worked. In a Kindle Direct Publishing pilot the firm mined public demand signals, produced titles with AI assistance, published them into its own distribution and measured the result, reporting click-through rates two to three times benchmark, ratings of 4.2 to 4.5 stars, and titles profitable within two to four months. The pilot is the prior evidence for a general pattern the firm can execute: **demand signal → AI production → measured library**. It is worth being precise about what it does and does not prove. It proves that the sequence can be executed, and that the measurement step is real rather than rhetorical. It does not prove that the sequence transfers to music, to a health-adjacent need, or to a market where the product must demonstrate an effect on the user's body rather than a click.

Figure 1.1 The firm's innovation engine, and the transfer this study examines



Sources: company/COMPANY_BACKGROUND.md §2 (asset inventory, KDP pilot outcomes); DESIGN_THINKING_FRAMEWORK.md framing update, 9 July 2026.
Catalogue and distribution figures are of 2022 vintage and are flagged for reconciliation in the company's own file.

Figure 1.1. *The firm's innovation engine as validated in the Kindle Direct Publishing pilot, and the transfer this dissertation examines.* The upper track is the validated instance: a demand signal mined from public sources, AI-assisted production, publication into owned distribution, and measurement against a benchmark, with the reported outcomes attached. The lower track is the proposed transfer to music for anxiety relief, with the two steps that do not carry over unchanged marked: the demand signal must be established rather than mined, and "measured" now means a measured effect on a person, not a measured click. Sources: `company/COMPANY_BACKGROUND.md` §2; `DESIGN_THINKING_FRAMEWORK.md` framing update.

1.2 Why this is corporate innovation and not a start-up

The distinction is not cosmetic; it changes what counts as an answer. A start-up would be asked whether the market is attractive. An established, profitable firm must additionally be asked two questions a start-up never faces: whether it holds assets that make this particular market attractive *to it*, and whether it can run an exploratory venture without the exploitation business smothering it.

The first question is the resource-based one. The firm's unfair advantages — an AI music production pipeline, an owned and controlled catalogue with rights, artist relationships, worldwide distribution and market knowledge in both China and the West — are precisely what a pure start-up in this category lacks. The endowments run in opposite directions: the most directly comparable competitor built its position on generative audio and has had to acquire artist relationships subsequently, through label deals, while this firm begins with the catalogue, the rights and the artists and without a generative audio product. Read through the resource-based view, the question is whether those assets are valuable, rare, inimitable and organisationally exploitable *in the new market* rather than in the old one (Barney, 1991), and the answer turns out to depend on an empirical fact about listeners rather than on an inventory of the firm.

The second question is the ambidexterity one. March (1991) framed the underlying trade-off: exploitation returns are proximate, certain and self-reinforcing, while exploration returns are distant, uncertain and easily starved by the very success of the core business. A firm earning $2.2 M of net income from distribution has an exceptionally strong incentive to spend its next increment of managerial attention on distribution. The standard structural remedy is a small, separate unit with its own resources and its own criteria, protected from the operating business's metrics (O'Reilly & Tushman, 2004, 2008), and that is the form the exploration reported here took: a ring-fenced budget, a small team, an explicit end date, and a decision rule written down in advance. Chapter 7 treats the organisational arrangement as a finding in its own right rather than as background.

In the vocabulary of corporate entrepreneurship this is internal corporate venturing: a new business developed inside an established firm, dependent for its survival on the strategic context that firm gives it rather than on the market alone (Burgelman, 1983a; Sharma & Chrisman, 1999). Positioned on the classical growth grid, the venture is an adjacent move rather than a leap: a new customer need (calm, health-adjacent) served by an adapted version of an existing product (music) built on existing assets — what the project's framing document calls "diversification-lite" (Ansoff, 1957), and what an innovation-portfolio view would classify as an adjacent rather than a core or transformational bet (Nagji & Tuff, 2012). On a horizon view, distribution and AI music creation are Horizon 1, the cash engine; the wellness product is Horizon 2; and a closed-loop "music as digital therapeutic" is Horizon 3, deliberately deferred (Baghai et al., 1999). These classifications are not decoration. They set the governance the venture should receive: Horizon 2 ventures are funded against learning milestones rather than revenue forecasts, which is exactly what the gate structure in chapter 4 implements.

1.3 Problem statement and research questions

Three facts make the problem non-trivial rather than merely commercial.

First, the need is large and structurally under-served. An estimated 359 million people had an anxiety disorder in 2021, and roughly 27.6 per cent of those needing treatment receive it (World Health Organization [WHO], 2023). Music listening reliably lowers state anxiety across independent meta-analyses — a Cochrane review of 26 perioperative trials (N = 2,051) found a reduction of 5.72 units on the state scale of the State–Trait Anxiety Inventory (95% CI [−7.27, −4.17]), performing comparably to a sedative in one trial (Bradt et al., 2013), and a 104-trial synthesis (N =

9,617) found effects on psychological (d = .545) and physiological (d = .380) stress markers, with heart rate the strongest physiological marker (de Witte et al., 2020) — but the underlying trial evidence is graded low quality, and it was produced in supervised clinical settings that do not resemble a phone in a bedroom.

Second, the incumbent solution set is commercially stressed rather than complacent. Calm's 2025 revenue is estimated at approximately $210 M, down 24 per cent year on year with roughly 3.5 M subscribers; Headspace's estimated revenue fell from about $348 M in 2024 to about $140 M in annual recurring revenue in 2025 alongside a 13 per cent workforce reduction; and industry-wide thirty-day retention for meditation applications is about 4.7 per cent, with an average of one to four lifetime sessions (Creswell & Goldberg, 2025). A category in which the two leaders are shrinking is not obviously a category to enter, and the retention figure is the single most important number in the business model of chapter 6.

Third, the evidence base the category leans on is thinner than its marketing suggests. Only Calm and Headspace have any randomised-trial coverage at all; half of Headspace's own trials disclose a conflict of interest; and no equivalent independent review exists for Endel, Brain.fm or any Chinese competitor (O'Daffer et al., 2022). A firm that can actually measure an effect therefore has an opening — but only if it measures.

The problem, then, is not whether a calming-audio application can be built. It plainly can, and one has been: the product described in chapter 6 is installable and was submitted for App Store review during the writing of this dissertation. The problem is whether a firm whose advantage is real music, and the rights to it, can win in a market whose users — in the state where they most need help — may not want real music at all. That tension is the spine of the dissertation, and it is what makes the case worth reporting: the evidence turned against the firm's own principal asset, and the process had to decide what to do about it.

Three research questions follow. The first two are the questions the firm's supervisor put to the design-thinking phase; the third is the question this dissertation adds, and the one to which chapter 7 is the answer.

Table 1.1. *The three research questions, the claims each rests on, and where each is estimated*

	Question	Claim to be established	Principal evidence	Where estimated
RQ1 · Need	Is there a sizeable, reachable, under-served need that the firm can address?	N1 a sizeable segment has frequent "need to calm down" moments; N2 current solutions under-serve specific outcomes; N3 the need is state-dependent (acute versus wind-down) and so is the content that serves it	Survey (quantitative); interviews and focus groups (qualitative); netnography	Survey blocks B–C; focus-group themes; interview synthesis (ch. 5)
RQ2 · Ability	Can the firm serve it with its assets, measurably, and with a viable business model?	A1 content preference by state; A2 one session moves state anxiety and heart-rate measures; A3 users connect a wearable and value the result; A4 someone pays; A5 adaptation rights are feasible	Pilot experiment; survey; landing page (behavioural); legal check	Pilot mixed model; survey blocks D–F; belief register (ch. 5–6)

	Question	Claim to be established	Principal evidence	Where estimated
RQ3 · Process	How should a firm of this type organise and decide such an exploration, and what did this case teach about that process?	P1 the organisational form that makes an honest stop possible; P2 the decision rule that survives contact with disconfirming evidence; P3 the capability the firm actually acquired	The project's own dated record: decision model, task division, decision log, artefacts	Chapter 7, discussed in chapter 8

RQ1 and RQ2 are transcribed from `research/METHODOLOGY_DESIGN.md` *§0. RQ3 is added by this dissertation.*

Six objectives follow directly: (i) establish whether the need exists and how it segments; (ii) establish which of the firm's assets actually transfer; (iii) measure whether one session changes anything, on self-report and on a device the user already owns; (iv) establish whether anyone pays, and at what price; (v) establish whether the rights permit the asset-leveraging version of the product; and (vi) end the phase with a documented decision rather than a pitch.

1.4 Scope and boundaries

Four boundaries were fixed before the work began, and each constrains what this dissertation can conclude.

Table 1.2. *The four boundaries fixed before the work began, and their consequences*

Boundary	Setting	Consequence
Positioning	General wellness	No diagnosis and no treatment claim anywhere in the product or in this document; the digital-therapeutic route is deferred until after funding
Go-to-market	Overseas first	The United States and the European Union for willingness to pay; China only as a low-cost probe, so Chinese market figures are context rather than a plan
Hardware	Oura and Apple Watch, Polar H10 as reference	The author owns an Oura ring; Apple Watch is the volume device; the chest strap is the calibration ground truth
Budget and time	$10–20k, ≤ 3 months	Ring-fenced for the design-thinking phase, ending in a decision memo

The wellness boundary means that no efficacy statement in this document is a medical claim. The overseas-first boundary means the China analysis in chapter 3 is context. The budget boundary means the efficacy pilot is 20 to 30 participants, which the methodology declares in advance to be a feasibility signal for self-report and under-powered for physiology (chapter 4.3.2) — a declaration made before the data rather than after. The phase boundary means that retention, the most load-bearing commercial variable in this category, cannot be tested at all inside the phase; it is carried as a watch item rather than as a gate. The amount of the ring-fenced budget actually spent by the time of writing is **[to confirm with the author]**.

Two further limits should be stated at the outset rather than discovered in chapter 8. The efficacy pilot had not been run at the time of writing, so every efficacy statement here is borrowed from meta-analyses conducted in other populations. And the survey and focus-group figures reported in chapter 5 and reused in chapters 6 to 8 are the output of the project's pre-registered analysis pipe-

line as published on the project site, not fieldwork; they are labelled as such wherever they appear, and fieldwork will replace them in a single pass.

1.5 Structure of the dissertation

Figure 1.2 Map of the dissertation

CHAPTER	WHAT IT ESTABLISHES	SERVES
1 · Introduction	Corporate-innovation framing, the engine, the three questions	framing
2 · Literature review	Five strands, and the gap none of them closes	RQ1 · RQ2 · RQ3
3 · Company and industry	VRIO, PESTEL, five forces, competitors, sizing, SWOT	RQ2
4 · Methodology	Mixed-methods design, the Bayesian layer, the gates, the pilot	all three
5 · Findings I — need and content	The state-dependence result, and the belief register's verdict	**RQ1**
6 · Findings II — concept and model	Concepts, canvases, price, unit economics, the built product	**RQ2**
7 · Findings III — the process the original contribution	How the exploration was organised, decided and audited	**RQ3**
8 · Discussion	Answers, theory, managerial implications, reflexivity, limits	all three
9 · Recommendations, roadmap	The decision memo, the risk register, the resource plan	decision
10 · Conclusion	What is answered, what is not, and what should be studied next	—

Shaded band: the findings chapters. Chapter 7 answers the research question this study adds to the two the firm's supervision set.

Figure 1.2. *Map of the dissertation.* Each chapter is shown with the research question it serves and the evidence it rests on. The three shaded chapters (5, 6, 7) are the findings; of these, chapter 7 answers the research question added by this study.

Chapter 2 reviews five literatures — corporate entrepreneurship and ambidexterity, innovation typologies, resource-based and dynamic-capability views, managing innovation under uncertainty, and design thinking as a corporate method — together with the domain evidence on music and autonomic arousal, and names the gap this study addresses. Chapter 3 analyses the company and the industry with the standard strategy toolkit. Chapter 4 sets out the methodology: the philosophy and the insider's position, the mixed-methods architecture, the instruments, the Bayesian decision layer with its pre-registered gates, and the efficacy pilot design. Chapter 5 reports the findings on need and content. Chapter 6 turns those findings into a concept, a business model and a built product. Chapter 7 — the original contribution — reports on the innovation process itself: how the exploration was organised, how each gate rule shaped a decision, what the discipline changed, and what capability the firm acquired. Chapter 8 discusses the answers to the three research questions and their theoretical and managerial implications, states the author's own influence on the work, and sets out the limitations. Chapter 9 gives the recommendation, the roadmap and the risk register; chapter 10 concludes.

2. Literature review

This chapter reviews the literatures on which the study draws and identifies the gap it addresses. It is organised as ten strands. The first eight are management literatures: corporate entrepreneurship (2.1), exploration and exploitation (2.2), innovation typologies (2.3), the resource-based and dynamic-capability views (2.4), the management of innovation under uncertainty (2.5), design thinking as a corporate innovation method (2.6), business-model innovation (2.7), and evidence-based decision-making (2.8). The ninth summarises the domain evidence on music, autonomic arousal and digital wellness products (2.9), which is context for the case rather than its contribution. The tenth states the gap and the contribution (2.10).

Each strand closes with a short paragraph headed *Implications for the case*, which connects the theory to TDMusic and to the adaptive-audio product the firm has been developing under the name Lilt. Those paragraphs draw only on material already documented in the project repository — the company's asset inventory, the design-thinking framework and its artefacts, the belief register and its pre-registered gates, and the verified research findings — so that the theoretical reading and the empirical record stay separable. Where the literature is invoked to license a later analytical move, the move is named, so that the reader can check in chapters 4 to 7 whether the licence was used as described.

Two conventions apply throughout. First, every work cited here has been verified against a digital object identifier, a publisher deposit, a national library or index record, or the publisher's own page; the verification trail is recorded in `dissertation/sources/verified.md`, and works that could not be verified are listed in `dissertation/sources/unverified.md` and are not cited. Second, where the physiology of relaxation is discussed, the direction is stated explicitly and consistently: relaxation *raises* heart-rate variability and *lowers* heart rate. The product and its claims are framed as general wellness throughout; no therapeutic or diagnostic claim is made or reviewed as though it were available to the firm.

2.1 Corporate entrepreneurship and intrapreneurship

The study's unit of analysis is not a start-up but an entrepreneurial episode inside an established, profitable firm, and the literature that describes such episodes is corporate entrepreneurship. Its empirical foundation is Burgelman's process research on internal corporate venturing. Burgelman (1983a) modelled internal corporate venturing in a diversified major firm as the interaction of two loops: a *definition* process in which technical and market possibilities are articulated at operational level, and an *impetus* process in which managerial championing converts a project into a resource-attracting venture. Around these sit *strategic context determination*, by which autonomous initiatives are retrospectively fitted into corporate strategy, and *structural context determination*, by which corporate management shapes the administrative arrangements that select among initiatives in the first place. Burgelman (1983b) drew the strategic-management consequence: firms generate both induced strategic behaviour, which fits the current concept of strategy, and autonomous strategic behaviour, which does not, and the managerial task is to keep the second alive long

enough to be assessed rather than filtering it out early. Burgelman (1984) then set out the design consequence, arranging corporate venturing forms — from a direct integration or new-product department at one extreme, through new-venture divisions and independent business units, to nurturing-and-contracting arrangements and complete spin-offs — on two dimensions: the venture's *strategic importance* to the corporation and its *operational relatedness* to existing capabilities.

The definitional literature that followed is concerned less with process than with scope. Guth and Ginsberg (1990), introducing the field's first dedicated special issue, defined corporate entrepreneurship as encompassing two distinct phenomena: the birth of new businesses within existing organisations, and the transformation or strategic renewal of those organisations. Sharma and Chrisman (1999) reconciled the competing definitions then in circulation and proposed a taxonomy that has largely held: corporate entrepreneurship comprises *corporate venturing* (internal, cooperative or external), *innovation*, and *strategic renewal*, with corporate venturing further split by whether the new business is created inside the firm's boundaries and by the degree of structural autonomy granted to it. Covin and Slevin (1991) supplied the firm-behaviour model that made entrepreneurial orientation measurable, and Ireland et al. (2009) later formalised corporate entrepreneurship *strategy* as the joint product of an entrepreneurial strategic vision, a pro-entrepreneurship organisational architecture, and entrepreneurial processes and behaviour. Alongside this academic line runs a practitioner tradition, established by Pinchot (1985), which coined "intrapreneuring" for the individual who behaves entrepreneurially without leaving the corporation, and by Kanter (1985), whose field research identified the structural supports — access to resources outside the normal budget cycle, protection from premature evaluation, and a clear route back for people whose venture fails — that established companies must supply if such individuals are to persist.

Three things in this literature matter for the present study. The first is that the field treats structural autonomy as a design variable, not a virtue: Burgelman's (1984) matrix explicitly reserves high autonomy for ventures that are strategically important *and* operationally unrelated, and prescribes tighter integration where operational relatedness is high. The second is that the same literature is consistently attentive to the filtering problem — autonomous initiatives die because the firm's existing structural context selects against them, not because they are refuted. The third is that corporate entrepreneurship is defined by its *organisational* location rather than by its novelty: the phenomenon of interest is what an established firm does with an initiative, not how novel the initiative is in the abstract.

Implications for the case. TDMusic is a profitable, mid-sized digital-content distribution firm — the company file records revenue of $10.5 million, net income of $2.2 million and 52 employees — whose validated operating engine runs from demand signal to AI-assisted production to a measured, distributed library, and which has already demonstrated that engine outside music in the KDP publishing pilot. The Lilt exploration is therefore internal corporate venturing in Burgelman's sense rather than either a diversification acquisition or a start-up: it is an autonomous initiative seeking strategic context, championed inside a firm whose induced strategic behaviour is distribution. On Burgelman's (1984) two dimensions, the venture is of moderate strategic importance and *high* operational relatedness — it proposes to run the firm's own engine over the firm's own catalogue and rights — which argues for a small, semi-autonomous team with privileged ac-

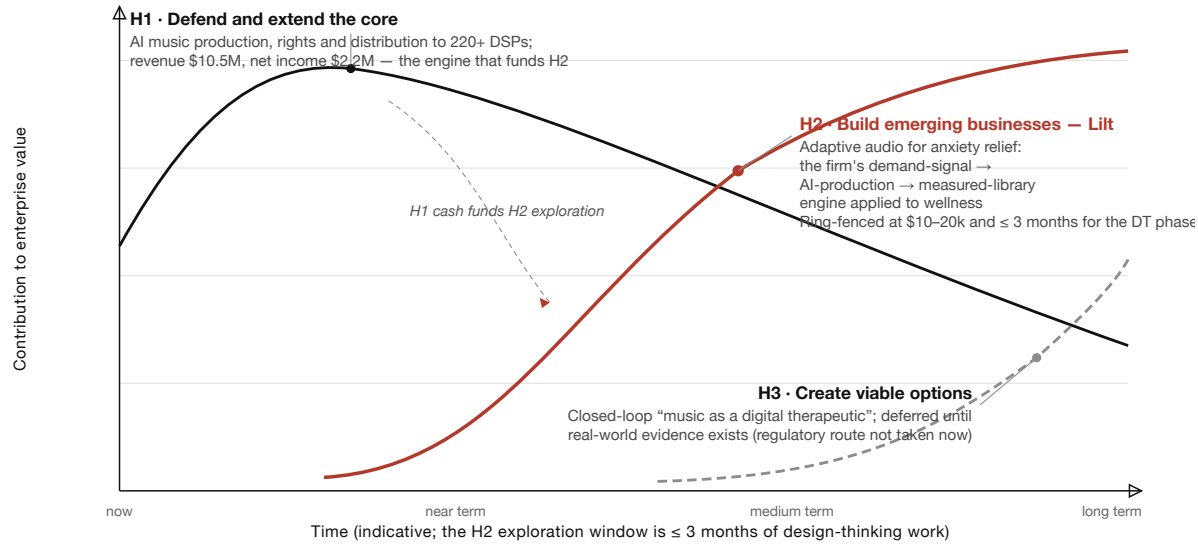
cess to the core assets rather than a separate new-venture division. That is what the design-thinking framework specifies: a small team, a ring-fenced budget of $10,000 to $20,000, and a phase of no more than three months. Kanter's (1985) protection-from-premature-evaluation condition is met not by shielding the venture from evidence but by fixing the evaluation rule in advance, which is the function the belief register performs.

2.2 Exploration, exploitation and ambidexterity

March (1991) supplied the tension that organises this chapter. Exploitation — refinement, efficiency, selection, execution — yields returns that are proximate in time, certain in magnitude and local in effect. Exploration — search, variation, experimentation, discovery — yields returns that are distant, uncertain and often captured by someone else. Because adaptive learning processes reward what has already worked, they tend to drive out exploration even when exploration would be optimal in the long run; a firm that learns quickly can converge prematurely on an inferior equilibrium. The problem is thus not that managers undervalue exploration in principle but that ordinary organisational learning suppresses it in practice.

Two families of solution followed. The structural family, developed by Tushman and O'Reilly (1996) and popularised by O'Reilly and Tushman (2004), separates exploration from exploitation into organisationally distinct units with different processes, structures and cultures, integrated only at senior level so that the exploratory unit can draw on corporate assets without inheriting the core business's metrics. O'Reilly and Tushman (2008) later reframed ambidexterity itself as a dynamic capability, arguing that the capacity to reallocate assets between exploitation and exploration is a senior-management competence rather than a structural fact, and that it rests on a clear strategic intent, an overarching identity, explicit senior-team ownership of the tension, and separate but aligned architectures. The contextual family, articulated by Gibson and Birkinshaw (2004), locates ambidexterity in the behavioural capacity of an entire business unit to pursue alignment and adaptability simultaneously; in their study, a context combining performance discipline and stretch with support and trust produced ambidexterity, which in turn mediated the relationship between context and business-unit performance. Raisch and Birkinshaw (2008) reviewed the field and organised it around recurring choices — differentiation versus integration, static structures versus dynamic sequencing, internal versus external sourcing of the exploratory activity — and Raisch et al. (2009) argued that these are not mutually exclusive alternatives but complementary mechanisms whose combination is the real research question.

Practitioner strategy supplies a complementary temporal frame. Baghai et al. (1999) proposed the Three Horizons: Horizon 1 defends and extends the core businesses that generate today's cash and earnings; Horizon 2 builds the emerging businesses that will become tomorrow's earnings and that consume, rather than generate, cash; Horizon 3 creates viable options — small, cheap positions that may or may not mature. The model's discipline is that all three must be staffed and funded concurrently, and that each is evaluated on different metrics; applying Horizon 1 profitability tests to a Horizon 2 business is the model's canonical failure mode, and is exactly the filtering problem Burgelman identified.



Horizons after Baghai, Coley and White (1999); horizon content from the firm's own framing document and company file.

Figure 2.1. *Three Horizons of growth applied to TDMusic.* Horizon 1 is the distribution and AI-production cash engine; Horizon 2 is the Lilt adaptive-audio wellness product, funded from Horizon 1 and ring-fenced in budget and time; Horizon 3 is a closed-loop digital therapeutic held as an option and not funded in this phase. Horizons after Baghai et al. (1999); horizon content from the firm's own framing document and company file.

Implications for the case. The firm's own framing document assigns AI music creation and distribution to Horizon 1, the music-for-wellbeing product to Horizon 2, and a closed-loop "music as digital therapeutic" to Horizon 3, and the company file states the funding logic plainly: Horizon 1 net income funds the Horizon 2 exploration. Figure 2.1 renders that assignment. The structural provision is modest by O'Reilly and Tushman's (2004) standards — a small separate team with a ring-fenced budget and timetable rather than a distinct business unit — which is proportionate to a 52-person firm but leaves the exploration exposed on the dimension Gibson and Birkinshaw (2004) emphasise: the same person is the founder, the champion and the researcher, so the behavioural context, not the organisational chart, is what has to carry the discipline. The methodology document makes this explicit, treating the small separate team and the ring-fenced budget as "the organisational condition for honest kill decisions". Read through March (1991), the risk is not that the firm will fail to explore — it has already built and submitted a product — but that it will explore *in the direction its existing competences make comfortable*, which is precisely what the real-music-versus-neutral- sound tension puts at stake.

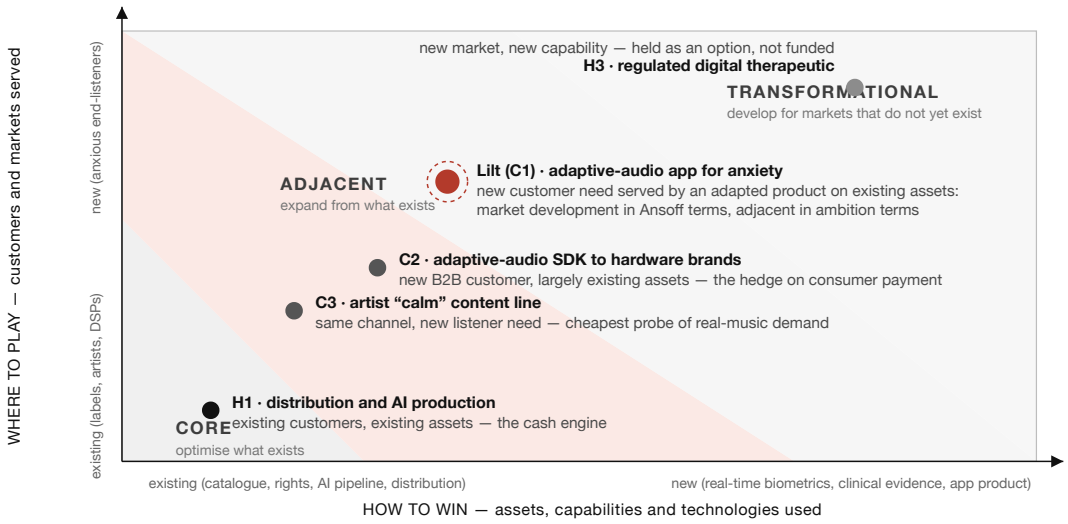
2.3 Innovation typologies: how much of a departure is this?

Typologies of innovation exist to answer a governance question: how far from the firm's existing knowledge and markets does a given initiative sit, and therefore what kind of management does it need? The oldest cut is incremental versus radical, but Garcia and Calantone (2002) showed in a systematic review that the terminology is used inconsistently across studies — "radical", "really

new" and "discontinuous" are applied to different constructs measured at different levels of analysis — and recommended that any classification state explicitly whether it is assessing novelty to the firm, to the market, or to the world, and on which of technology and marketing dimensions. Abernathy and Clark (1985) had already provided a two-dimensional answer, classifying innovations by their capacity to conserve or destroy existing *technological* competence and existing *market/customer* linkages, and naming the quadrants architectural, niche-creation, regular and revolutionary.

Henderson and Clark (1990) sharpened the technological side into the distinction that matters most here. Separating knowledge of components from knowledge of the *architecture* that links them, they showed that innovations which leave components essentially unchanged while reconfiguring the way those components are connected are systematically destructive of established firms' capability — precisely because architectural knowledge, unlike component knowledge, becomes embedded in a firm's information filters, communication channels and problem-solving routines, and is therefore invisible to the people who hold it. Established firms fail at architectural innovation not through inattention but because they process new information through an architecture that no longer applies.

Christensen and Bower (1996) and Christensen (1997) supplied the demand-side counterpart: firms that listen attentively to their most profitable existing customers rationally decline to fund technologies that initially underperform on the attributes those customers value, and are displaced when the trajectory of the new technology intersects mainstream requirements. Christensen et al. (2015) subsequently restated the theory's boundary conditions, insisting that disruption is a *process* originating in low-end or new-market footholds and that the label is routinely misapplied to any successful new entrant. Nagji and Tuff (2012) translated the same intuition into a portfolio instrument, the innovation ambition matrix, which arrays initiatives by how novel the assets and capabilities are ("how to win") and how novel the customers and markets are ("where to play"), and distinguishes core, adjacent and transformational ambition; their survey evidence suggested that firms with a roughly 70/20/10 allocation across the three outperformed peers, and, more useful for a single case, that the three require different funding, metrics and talent. Ansoff (1957) had supplied the underlying growth vectors — market penetration, market development, product development and diversification — which remain the cleanest way to say what is new about a given move.



Matrix after Nagji and Tuff (2012); axis reading after Ansoff (1957). Placements are the author's, from the firm's asset inventory and concept set.

Figure 2.2. *Innovation ambition matrix with the firm's options placed.* The distribution business sits in the core; the artist "calm" content line and the adaptive-audio SDK sit in the near-adjacent band; Lilt sits in the adjacent band as a new customer need served by an adapted product on existing assets; a regulated digital therapeutic sits in the transformational corner and is held as an unfunded option. Matrix after Nagji and Tuff (2012); axis reading after Ansoff (1957).

Implications for the case. In Ansoff's (1957) terms Lilt is market development shading into diversification: the customer need is new (personal anxiety relief rather than catalogue consumption by platforms and their listeners), the product is an adaptation of an existing one, and the assets are largely existing. In Nagji and Tuff's (2012) terms it is adjacent rather than transformational, which is how Figure 2.2 places it — and that placement carries the governance implication that the venture should be resourced and measured differently from the Horizon 1 business but does not warrant transformational-scale investment, which is consistent with the phase budget of $10,000 to $20,000 recorded in the framework. The sharpest reading, however, is Henderson and Clark's (1990). The firm's *component* knowledge transfers well: AI-assisted production of calm arrangements, rights and royalty administration, and distribution to more than 220 digital service providers are all directly reusable, and the company file names them as such. What does not transfer is the *architecture*: a consumer subscription relationship, a fifteen-minute listening session with a measured before-and-after, and a recommendation loop that answers to an individual's arousal state rather than to a platform's playlist economics. The firm's existing filters are tuned to per-track revenue across a catalogue, not to whether one person is calmer after one session. Christensen et al.'s (2015) discipline is worth applying in the negative: nothing in the repository supports calling Lilt disruptive, because it targets neither a low-end foothold in an existing market nor a population of established non-consumers whose incumbents would rationally ignore them, and the chapter therefore does not use the term.

2.4 Resources, capabilities and the dynamic-capability view

If corporate entrepreneurship asks what the firm is doing, the resource-based view asks what entitles it to succeed. Wernerfelt (1984) reframed the firm as a bundle of resources rather than a set of product-market positions, and Barney (1991) specified the conditions under which such a bundle can yield sustained competitive advantage: the resource must be valuable, rare, imperfectly imitable and without strategically equivalent substitutes. Barney (1995) restated the framework in the form now used in practice, adding the question of whether the firm is *organised* to exploit the resource, which converts the schema into the VRIO test applied in chapter 3.

The static character of that account drew two responses. Teece et al. (1997) argued that in environments of rapid change advantage rests less on the resource stock than on *dynamic capabilities* — the firm's ability to integrate, build and reconfigure internal and external competences — and located those capabilities in processes, positions and paths rather than in tradable assets. Teece (2007) then disaggregated dynamic capabilities into three clusters with identifiable microfoundations: *sensing* opportunities and threats, which is an analytical and creative activity resting on search and interpretive routines; *seizing* them, which requires committing resources through business-model design, boundary decisions and the avoidance of decision biases; and *reconfiguring* — managing threats and continuing to realign assets as the opportunity matures. Eisenhardt and Martin (2000) offered the principal caveat, arguing that dynamic capabilities are identifiable, partly idiosyncratic and partly common best practice, and that in high-velocity markets they take the form of simple rules and real-time information rather than elaborate analytic routines — so that the capability's value lies in the pattern of resource reconfiguration it produces, not in the capability itself. Danneels (2002) supplied the bridge to the present case, showing that product innovation both draws on and builds two kinds of competence — *technological* competence in making a thing, and *customer* competence in knowing and reaching the people who use it — and that firms extend their scope by leveraging one while acquiring the other, a process he called second-order competence.

Implications for the case. The company file's asset inventory maps directly onto this vocabulary. The rights and catalogue position — of the order of 100,000 songs, the large majority owned or directly licensed, with rights administration already in place — is the one asset with a plausible claim to Barney's (1991) inimitability condition, because rights are legally exclusive; but its value for this venture turns entirely on whether per-track adaptation and derivative rights actually exist, which the register carries as assumption A5 and which the pending legal check is designed to settle. The AI production pipeline and the published recommendation algorithm are valuable and organised for use, but the repository holds no benchmark against competing systems, so rarity and inimitability cannot be asserted for them. Teece's (2007) triad supplies the architecture of the argument this dissertation makes about process: the design-thinking phase is the firm's *sensing* routine, the pre-registered gates are its *seizing* discipline — they are exactly the mechanism Teece identifies for committing resources while guarding against decision bias — and the reconfiguring options are the SDK route to hardware partners, the rights audit, and a library that becomes measured and self-generating as sessions accumulate. Danneels (2002) names the gap most precisely: the firm has deep technological competence in producing and distributing music and no estab-

lished customer competence for an anxious individual listener, and the design-thinking phase is, in his terms, an attempt to build that second-order competence cheaply before committing to it.

2.5 Managing innovation under uncertainty

Knight (1921) drew the distinction on which this whole strand rests: *risk*, where outcomes are unknown but their probability distribution is knowable and therefore insurable, and *uncertainty*, where the distribution itself is unknown. Profit, in Knight's account, is the return to bearing the second, not the first. The practical difficulty is that firms routinely apply risk instruments to uncertainty, producing forecasts whose precision is unwarranted.

Courtney et al. (1997) turned the distinction into a management tool by grading *residual* uncertainty — what remains after the best available analysis — into four levels: a clear-enough future, in which a single forecast is precise enough to act on; alternate futures, a small number of discrete outcomes to which probabilities can be attached; a range of futures, bounded but without natural discrete scenarios; and true ambiguity, where even a range cannot be constructed. Each level admits different tools and different strategic postures — shaping the market, adapting to it, or reserving the right to play — and different moves, from big bets through options to no-regrets actions. McGrath and MacMillan (1995) supplied the planning technique appropriate to the middle levels: discovery-driven planning inverts the conventional sequence by starting from the required outcome, deriving the operating performance the venture would have to achieve, listing every assumption that derivation depends on, and scheduling milestones so that each one tests the assumptions with the greatest leverage before further money is committed. McGrath (1999) then made the option logic explicit, arguing that entrepreneurial failure is manageable when ventures are framed as options — positions taken cheaply and abandoned quickly when the evidence disconfirms them — so that failure produces learning at bounded cost; Bowman and Hurry (1993) had earlier argued that firms hold bundles of "shadow options" in their existing resources, exercised incrementally as uncertainty resolves.

Two process traditions institutionalise these ideas. Cooper (1990) introduced the stage-gate system, in which development proceeds through defined stages separated by gates at which pre-specified deliverables are judged against pre-specified criteria and a go/kill/hold/recycle decision is made; Cooper (2008) updated the model for faster, more flexible and more open innovation, and Cooper and Sommer (2016) documented the agile hybrid in which sprint-based development is nested inside the gate structure. The lean-startup tradition works from the other end. Ries (2011) reframed the new venture as an experiment, proposing validated learning, the build–measure–learn loop, the minimum viable product, and *innovation accounting* — a discipline of measuring progress by movement in leading indicators about customers rather than by output. Eisenmann et al. (2011) codified the approach for teaching as hypothesis-driven entrepreneurship, and Blank (2013) argued that customer development plus agile engineering plus business-model hypotheses had become a general management method rather than a start-up technique. Sarasvathy (2001) offered the theoretical alternative of effectuation, in which entrepreneurs begin from available means and an affordable-loss threshold rather than from pre-specified goals and expected return,

and Sull (2004) described the resulting practice as disciplined entrepreneurship: a sequence of cheap tests attached to explicit decision rules.

The most useful recent evidence is experimental. Kerr et al. (2014) modelled entrepreneurship itself as a sequence of experiments whose value lies in cheap early failure. Camuffo et al. (2020) ran a randomised controlled trial with 116 Italian start-ups, training the treatment group to frame ideas as theories and test them scientifically; treated firms generated more revenue and — the finding that matters most here — were *more* likely to terminate their initial idea. Teaching firms to test rigorously does not make them more persistent; it makes them abandon bad ideas sooner. That is the empirical case for the apparatus this dissertation applies.

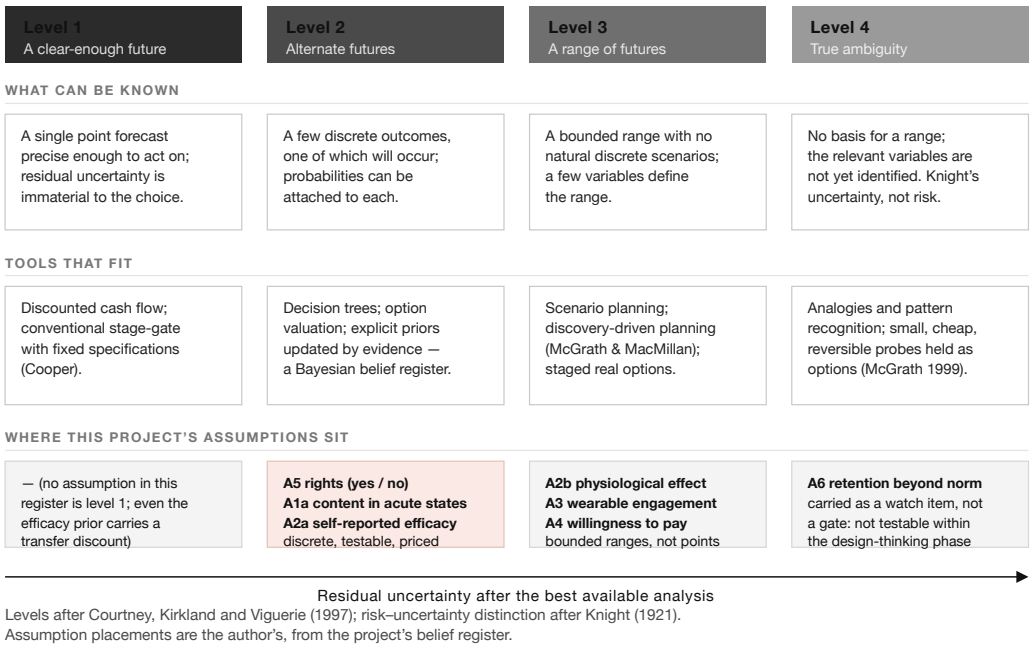


Figure 2.3. *Levels of residual uncertainty and the matching decision tools, with the project's own assumptions placed.* Levels after Courtney et al. (1997); the risk–uncertainty distinction after Knight (1921); assumption placements by the author from the project's belief register.

Implications for the case. Figure 2.3 places the register's assumptions on Courtney et al.'s (1997) scale. None sits at level 1: even the efficacy belief, which has meta-analytic support, carries a transfer discount because the stimulus and setting differ from the trials. The rights assumption (A5) is a clean level-2 question with two discrete answers, which is why the methodology document sequences the legal check first — it is the cheapest and most decisive purchase of information, and its pre-registered likelihood ratios (8.0 if positive, 0.15 if negative) are the largest in the register. Willingness to pay and engagement sit at level 3, where bounded ranges rather than point forecasts are the honest output; retention (A6) is close to level 4 and is therefore carried as a watch item rather than a gate, because it cannot be tested within a design-thinking phase. The project's method is thus a hybrid of the two process traditions the literature offers: stage-gate in its structure — pre-specified deliverables judged against pre-specified criteria, with an explicit kill option (Cooper, 1990, 2008) — and discovery-driven in its content, since the assumption checklist

and the milestone sequencing are precisely McGrath and MacMillan's (1995) prescription, and the ordering of tests by expected information per unit of cost is the real-options reasoning of Bowman and Hurry (1993) and McGrath (1999). Camuffo et al.'s (2020) result reframes what a pivot would mean in this case: if the discipline works, the most likely visible effect is not a faster launch but an earlier and better-founded change of direction.

2.6 Design thinking as a corporate innovation method

Design thinking supplies the sequence and the divergence–convergence discipline of the exploration, and is therefore reviewed as a method rather than as an epistemology. Brown (2008) presented it as a human-centred process combining desirability, feasibility and viability, moving through overlapping spaces of inspiration, ideation and implementation, and emphasised early, cheap prototyping and direct observation of users. Two process codifications are in general use: the five modes of the Stanford d.school — empathise, define, ideate, prototype, test (Hasso Plattner Institute of Design at Stanford, 2018) — and the Design Council's Double Diamond, whose contribution is the explicit alternation between divergent and convergent work across a problem space and a solution space (Design Council, n.d.).

Liedtka (2015) made the claim that gives design thinking a defensible role in a decision-heavy study. Rather than arguing that designers are more creative, she proposed that design thinking's tools function as a *social technology* that counteracts specific, documented cognitive biases in innovation decisions: direct customer observation and journey mapping counter projection bias and egocentric empathy gaps; explicit ideation of multiple alternatives counters hypothesis-confirmation bias; assumption surfacing and rapid, low-cost experimentation counter planning fallacy and endowment effects. Liedtka (2018) reported field evidence from organisational implementations consistent with that mechanism. The bias-reduction claim is important here because it is the same claim the Bayesian layer in section 2.8 makes, by a more measurable mechanism, and the two can therefore be compared rather than merely stacked.

The critical literature is substantial and should be represented honestly. Kimbell (2011) argued that "design thinking" incoherently bundles a cognitive style, a general theory of design and an organisational resource, and that its accounts privilege the individual designer's cognition over the situated, material practices through which design actually happens. Carlgren et al. (2016) found, across five large firms, that the concept is enacted very differently in different organisations and proposed a framework of five themes — user focus, problem framing, visualisation, experimentation and diversity — each with associated mindsets, practices and techniques, precisely so that studies can say which version they mean. Micheli et al. (2019) reviewed the field, catalogued its attributes and tools, and identified the weakness most relevant here: the evidence linking design-thinking practice to innovation outcomes is thin, and the construct is frequently measured by the presence of its artefacts. Nakata and Hwang (2020) tested the composition question empirically and found the relationship between design thinking and new-product performance to be contingent rather than uniformly positive. Iskander (2018) made the political criticism: because design

thinking's empathy work is bounded by the sponsor's frame and its convergence is controlled by the facilitator, it can systematically reproduce the status quo while appearing to challenge it.

Within the define phase, three instruments recur in the corporate literature and in this project. Christensen et al. (2016) set out jobs-to-be-done, which specifies the progress a customer is trying to make in a circumstance rather than the attributes of a product. Ulwick (2002) operationalised the same idea as outcome-driven innovation, in which customers rate desired outcomes for importance and current satisfaction and the gap between the two identifies under-served opportunities. Kano et al. (1984) classified quality attributes as must-be, one-dimensional, attractive, indifferent or reverse, distinguishing features whose absence causes dissatisfaction from features whose presence delights.

Implications for the case. The framework document adopts the d.school modes paced by the Double Diamond, and the project's define-phase artefacts are exactly the instruments named above: personas, jobs-to-be-done statements, a point-of-view statement, how-might-we questions, a weighted territory-scoring matrix, plus a survey instrument carrying outcome-driven importance–satisfaction items and Kano functional/dysfunctional pairs. Micheli et al.'s (2019) criticism applies directly and is accepted rather than deflected: those artefacts are *outputs of a process*, not findings, and the framework document itself writes several of them as worked examples ("E.g. …") rather than as products of coding, so chapters 5 and 6 report them as artefacts and separate them from results. Iskander's (2018) critique bites harder than usual in this case, because the sponsor of the empathy work is also the owner of the asset under test: three of three acute-anxiety interviewees rejected melody, lyrics and beat and asked instead for a featureless "soft wall" of sound, which is a finding directly against the firm's catalogue advantage. A design process whose convergence is controlled by the facilitator can absorb that finding as an edge case. What prevents it here is not the method but the constraint imposed on it — the belief register was written before the data, and the acute-state content assumption carries a pre-registered disconfirming likelihood ratio. Liedtka's (2015) bias-reduction claim is therefore not taken on trust in this study; it is given a mechanism that can be audited.

2.7 Business-model innovation

A venture of this kind must decide not only what to build but how value will be created, delivered and captured. Teece (2010) defined a business model as the design or architecture of those three activities, argued that it is logically distinct from strategy, and made the point most relevant here: technology has no value in itself, so a technological capability must be paired with a commercially viable model, and the choice of model is itself an act of design under uncertainty that must often be refined by trial. Chesbrough (2010) addressed the barriers, arguing that established firms find business-model change harder than technology change because new models conflict with existing asset configurations and processes and because managers' cognition is anchored on the logic that has worked; he prescribed experimentation with low-cost probes, effectual reasoning, and explicit leadership of the change. Zott et al. (2011) reviewed the field and identified three largely separate research streams — e-business, strategy and firm performance, and innovation and tech-

nology management — converging on a view of the business model as a boundary-spanning *activity system* that is a new unit of analysis rather than a synonym for revenue model.

The practitioner canon supplies the instruments. Osterwalder and Pigneur (2010) codified the business model canvas, with nine building blocks running from customer segments and value propositions through channels, relationships, revenue streams, key resources, activities and partnerships to cost structure. Osterwalder et al. (2014) added the value proposition canvas, which pairs a customer profile of jobs, pains and gains against a value map of products, pain relievers and gain creators, and insists on explicit fit. Bland and Osterwalder (2019) then supplied the testing library that connects those canvases to evidence, organising experiments by cost, evidence strength and the stage of the idea.

Implications for the case. Two business models are live in the repository and they differ in more than pricing. The C1 route is a business-to-consumer subscription in which the firm holds the end-customer relationship, which is a genuinely new activity system for a company whose current model captures value through per-track royalties across many distribution endpoints; on Chesbrough's (2010) account this is where the resistance should be expected, because the new model conflicts with the configuration that produces the existing $10.5 million of revenue. The C2 route licenses an adaptive-audio SDK and catalogue to hardware and wellness brands, which is closer to the firm's present logic — it monetises rights and production capability through partners, and the company file already records Peloton, Tesla and Virgin Airlines as distribution endpoints, that is, business-to-business-to-consumer precedents already in house. The register treats the two as a portfolio rather than a ranking: C2 is explicitly described as a hedge against consumer-payment friction, C3 (an artist-branded calm content line through existing distribution) as the cheapest probe of demand for real-music calm, and C1 as the asset-leverage bet. That is Teece's (2010) point about model choice being an act of design under uncertainty, and it is why the gates select which concept to fund rather than only whether to fund. The load-bearing variable in either model is retention, which the domain evidence in 2.9 shows to be the category's structural weakness.

2.8 Evidence-based decision-making

The final management strand concerns how evidence should be converted into a decision. Rousseau (2006) argued that management, unlike medicine, lacks a systematic practice of using the best available evidence, and that closing the research–practice gap requires both better evidence and decision practices that actually consult it; Pfeffer and Sutton (2006) documented the substitutes managers use instead — casual benchmarking, unexamined habit, and ideology. The behavioural literature explains why. Tversky and Kahneman (1974) established that judgement under uncertainty relies on heuristics that produce systematic error; Lovallo and Kahneman (2003) showed how the resulting optimism bias and inside view lead executives to overestimate benefits and underestimate costs of their own initiatives, and recommended reference-class forecasting as the corrective; Kahneman et al. (2011) turned the corrective into a governance instrument — a checklist of questions a decision-maker should ask about *the process that produced a*

recommendation, distinguishing errors in the proposer's analysis from errors the reviewer can detect.

The methodological literature this project borrows most directly comes from political science. Fairfield and Charman (2017) set out explicit Bayesian analysis for process tracing: rival hypotheses are stated, priors are assigned and justified, each piece of evidence is assessed for how much more likely it is under one hypothesis than another — a likelihood ratio — and the posterior is computed in odds form, with the authors emphasising that the discipline's value lies in making the inferential weight of each item explicit and in guarding against overweighting evidence that is only weakly diagnostic. Humphreys and Jacobs (2015) provided the integrative counterpart, showing how qualitative and quantitative evidence can be combined within a single Bayesian framework so that the two are weighted by their diagnostic value rather than by disciplinary preference. Where evidence quality itself must be graded, the GRADE Working Group's approach (Guyatt et al., 2008) supplies a transparent scheme for rating certainty in a body of evidence and the strength of the recommendation that follows.

Two supporting literatures complete the apparatus. Pre-registration — fixing hypotheses, analyses and decision rules before the data are seen — was argued by Nosek et al. (2018) to be the clearest available separation of hypothesis-generating from hypothesis-testing work; Banks et al. (2019) set out what open-science practices mean for management and organisational research, and Toth et al. (2021) evaluated preregistration as a reporting method and documented how often registered plans and published analyses diverge, which is a caution rather than a refutation. Calibration research shows that probabilistic judgement is trainable: Mellers et al. (2014) found in a large geopolitical forecasting tournament that probability training, team collaboration and tracking of the best forecasters each improved accuracy, and Tetlock and Gardner (2015) drew out the practical implications — granular probability estimates, frequent small updates, and scoring against outcomes. Spetzler et al. (2016) place the same ideas in a corporate decision-quality frame, in which a decision is judged by the quality of its frame, alternatives, information, values and reasoning rather than by its outcome.

Implications for the case. The project's decision layer implements this literature almost line for line. Each critical assumption carries a prior with a written rationale and a source; each piece of evidence is converted into a likelihood ratio using a published verbal-strength rubric; likelihood ratios are then shrunk in log space by an evidence-quality coefficient — a GRADE-style move, with independent meta-analyses at full weight, industry-funded or single studies at roughly half, and small self-selected qualitative samples lower still — and correlated items are discounted before multiplication. The decision itself is a comparison of posteriors against thresholds fixed in advance: the register's proceed rule requires the efficacy pilot to have run and sets its thresholds deliberately above what the meta-analytic prior alone reaches, so that a proceed decision cannot be cleared by literature review. There are separate content and form pivot rules, and an explicit kill rule under which a negative result is recorded as a valid outcome. Qualitative evidence is allowed to move a belief but not, on its own, to cross a gate — the formal answer to the question of how much three interviews should count, and a direct application of Humphreys and Jacobs's (2015) weighting logic. The apparatus also has costs that the literature names and this study must report:

the likelihood ratios are subjective, the multiplication assumes conditional independence that the shrinkage only partly repairs, and Toth et al.'s (2021) finding that registered plans and executed analyses often diverge is an uncomfortable but relevant caution for a single-researcher study in which the analyst is also the founder.

2.9 The domain: music, autonomic arousal and digital wellness products

The domain evidence is context rather than contribution, and is summarised here at the level of what it licenses. Its detailed verification, including the fact-check verdicts on individual numeric claims, is held in the repository's research findings.

Music and self-reported anxiety. Multiple independent meta-analyses find a genuine anti-anxiety effect of music listening, with pooled effect sizes ranging roughly from d = −0.30 to d = −0.97 depending on population, comparator and outcome type. The most rigorous single synthesis is a Cochrane review of 26 perioperative trials (N = 2,051), which found a reduction of 5.72 STAI-S units, 95% CI [−7.27, −4.17], and an effect comparable to midazolam in one N = 327 trial, but rated the underlying evidence "low" quality because of the absence of blinding (Bradt et al., 2013). A broader synthesis of 104 randomised trials (N = 9,617) reported d = .545 on psychological and d = .380 on physiological stress outcomes (de Witte et al., 2020). Harney et al. (2023) report d = −0.77 across 21 controlled studies, rising to −0.97 in a low-risk-of-bias subset. Pelletier (2004) reached a directionally consistent d = .67 in the field's earliest quantitative synthesis. Two counterweights matter. Panteleeva et al. (2018) found a significant self-report effect (d = −.30) but a non-significant psychophysiological one in non-clinical samples — self-report and physiology need not move together. And Lassner et al. (2025), in the best available head-to-head synthesis across delivery modes, found similar effect sizes for active music therapy, receptive therapy and passive "music medicine" (SMD = 0.47, k = 23, n = 1,084) but rated the evidence very low quality and found the effect *not maintained at follow-up*.

Autonomic arousal. The physiological direction is consistent and mechanistically well supported. Slow-tempo music of roughly 60 to 80 beats per minute increases parasympathetic (vagal) tone — that is, it *raises* heart-rate variability — and *lowers* heart rate, blood pressure and respiratory rate, while faster, more stimulating music shifts the balance toward sympathetic dominance; tempo matters more than genre (Bernardi et al., 2006), a relationship corroborated by a recent review (Mitrovic & Paladin, 2026). The closest supporting evidence for a closed-loop design is indirect: a meta-analysis of heart-rate-variability biofeedback found a large effect on self-reported stress and anxiety (Hedges' g = 0.81 pre–post; g = 0.83 versus control; Goessl et al., 2017), but those protocols are breathing-paced rather than music-driven, and the generalisation gap should be stated rather than assumed.

Generative and adaptive audio. Peer-reviewed evidence for commercial generative soundscapes is close to absent. The single located study linked to the leading generative-audio product measured *focus* rather than anxiety, heart-rate variability or cortisol, was funded by the companies involved and used authors employed by one of them (Haruvi et al., 2022). The nearest mechanistically specific positive result concerns amplitude-modulated music and attention, and carries the

same industry-funding caveat (Woods et al., 2024). No controlled trial was located that compares real, personally meaningful music against engineered neutral sound for anxiety.

Digital wellness products as businesses. Independent efficacy evidence across the commercial category is thin: a systematic review of randomised trials of the two market leaders found limited coverage and a substantial share of trials carrying conflicts of interest (O'Daffer et al., 2022). The category's structural problem is retention rather than acquisition: Creswell and Goldberg (2025) report that roughly 4.7 per cent of meditation-app users are still using the app after 30 days, with an average of one to four lifetime sessions, even though the two leaders hold the great majority of category monthly active users. Prevalence and willingness-to-pay context is regionally uneven — China's national epidemiological survey reports markedly lower anxiety prevalence than United States survey estimates (Huang et al., 2019), a gap the repository's verified market file attributes to under-diagnosis, stigma and instrument differences rather than to a true difference in distress, and the same file records structurally lower willingness to pay in the Chinese market. Regulatory framing is a constraint on claims rather than on the product: general-wellness positioning permits language about relaxation and stress management and excludes claims to treat a named disorder, and the repository records two prominent cases in which regulatory clearance did not deliver commercial viability.

Implications for the case. Four design consequences follow, all of which the repository has already encoded. First, the primary outcome for the planned efficacy pilot is self-reported state anxiety, with heart rate as the accurate physiological co-measure and heart-rate variability as a noisier secondary — a split that follows directly from Panteleeva et al.'s (2018) divergence between self-report and physiology and from the finding that heart rate is the strongest of the physiological markers. This is why the register separates the self-report belief (A2a, posterior ≈ 0.85 on literature alone) from the physiological one (A2b, ≈ 0.56), and why the two carry different gate thresholds. Second, the pilot's two hypotheses are stated with the direction fixed: anxiety scores should fall further under the adaptive condition than under an active control, and heart-rate variability should *rise* while heart rate falls. Third, the literature does not answer the question on which the firm's asset advantage depends. There is no controlled comparison of real music against engineered neutral sound for acute anxiety, which is precisely the comparison the register's A1a and A1b assumptions encode and the pilot's T1/T2 conditions are designed to run. Fourth, the retention evidence sets the bar for the business model: the register carries retention as A6 with a prior of 0.25 against a category base rate of about 5 per cent at day 30, and explicitly declines to make it a gate because it cannot be tested inside this phase. All product language stays within general-wellness framing.

2.10 Gap and contribution

Laying the strands side by side reveals four gaps, of which the third and fourth are the ones this study is positioned to address.

The efficacy literature does not cover the product. The strong evidence is for music listening in supervised, largely perioperative settings, and for breathing-paced heart-rate-variability biofeed-

back. There is no meta-analysis, and no adequately powered trial, of biometric-adaptive audio delivered by a consumer application (Bradt et al., 2013; Goessl et al., 2017; Haruvi et al., 2022).

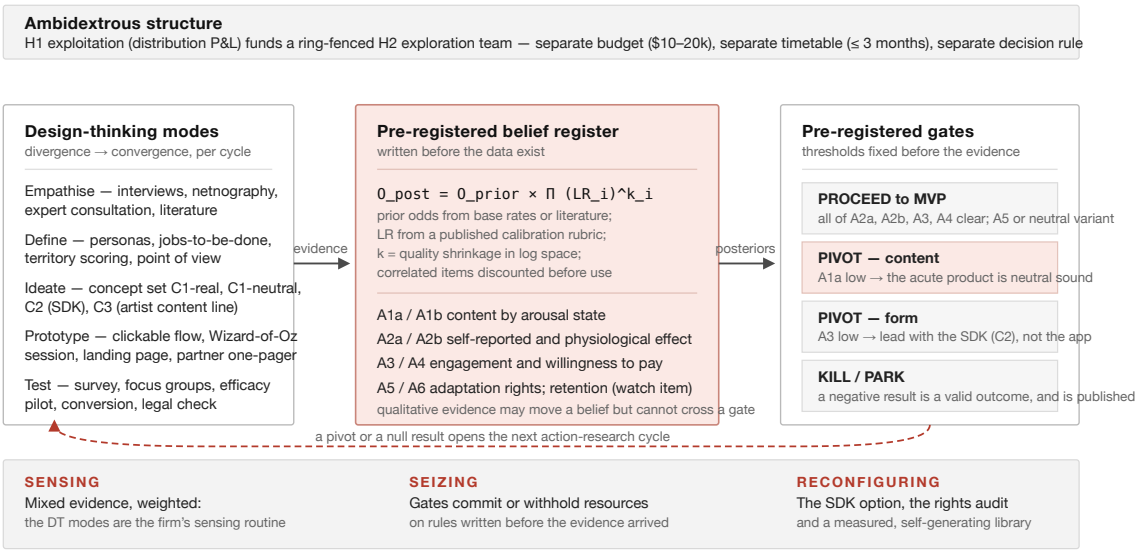
The content question is unasked. No controlled anxiety trial directly compares real, personally meaningful music with engineered neutral sound — the comparison on which this firm's asset advantage rests (Harney et al., 2023; Lassner et al., 2025).

Corporate-innovation research rarely shows its decision arithmetic. The ambidexterity, stage-gate and discovery-driven literatures agree that a firm should commit and withdraw resources on evidence (March, 1991; Cooper, 1990; McGrath & MacMillan, 1995; O'Reilly & Tushman, 2008), and the experimental evidence indicates that firms trained to test rigorously terminate weak ideas sooner (Camuffo et al., 2020). But published corporate cases seldom show the belief, the threshold and the update. The gate is reported as an outcome, not as an inference. Where an explicit apparatus exists for showing that inference — Bayesian process tracing with stated priors, likelihood ratios and quality shrinkage (Fairfield & Charman, 2017; Humphreys & Jacobs, 2015; Guyatt et al., 2008) — it has been developed for scholarly case inference in political science, not for resource-commitment decisions inside a firm.

Design-thinking write-ups rarely distinguish artefact from result. The critical literature makes the point in several registers: the construct is enacted differently in different firms (Carlgren et al., 2016), the link to outcomes is under-evidenced and often measured by the presence of artefacts (Micheli et al., 2019), its effect is contingent (Nakata & Hwang, 2020), and its convergence can quietly reproduce the sponsor's prior position (Iskander, 2018). Liedtka's (2015) bias-reduction claim is the field's most testable proposition, and it is rarely given a mechanism that an outside reader could audit.

The contribution is correspondingly specific. This dissertation documents, end to end and in a live case, an integrated process in which design thinking generates and shapes the options and a preregistered Bayesian belief register decides among them: priors with written rationales, likelihood ratios drawn from a published calibration rubric, GRADE-style quality shrinkage applied in log space, correlation discounting, and gates expressed as posterior thresholds fixed before the evidence arrived. Read against the management literature, this is an operationalisation of discovery-driven planning (McGrath & MacMillan, 1995) in which the assumption checklist becomes a probabilistic register and the milestones become likelihood ratios priced in advance; read against Teece (2007), it is a documented account of sensing and seizing routines in a firm that has neither before; and read against Liedtka (2015), it supplies design thinking's bias-reduction claim with an auditable mechanism rather than an assertion.

The case is unusually well suited to that demonstration for one reason: the evidence turned against the firm's own asset. The founding premise was that a music company's catalogue and rights were the unfair advantage in adaptive audio for anxiety; the empathy work and then the survey pointed the other way for acute states, and the belief register recorded the movement rather than absorbing it. A discipline for deciding evidence by evidence is worth little when the evidence flatters the incumbent's assets. Its value is visible only in the case where it does not, and that is the case documented here.



Synthesis by the author. Modes after Brown (2008) and Liedtka (2015); update rule after Fairfield and Charman (2017) and Humphreys and Jacobs (2015); gate logic after Cooper (1990) and McGrath and MacMillan (1995); capability vocabulary after Teece (2007); structure after O'Reilly and Tushman (2008).

Figure 2.4. *Conceptual framework of the dissertation.* Design-thinking modes generate evidence that enters a pre-registered belief register of priors, likelihood ratios and quality shrinkage; posteriors meet gates that return proceed, pivot or kill; the sequence is read as sensing, seizing and reconfiguring inside an ambidextrous structure funded by the core business. Synthesis by the author, drawing on Brown (2008), Liedtka (2015), Cooper (1990), McGrath and MacMillan (1995), Fairfield and Charman (2017), Humphreys and Jacobs (2015), Teece (2007) and O'Reilly and Tushman (2008).

Figure 2.4 states the framework the remaining chapters follow. Chapter 4 specifies the belief register, the calibration rubric and the gates as method; chapters 5 and 6 report the evidence and the concept work that passed through them; chapter 7 answers the process question by describing how the exploration was organised and what the discipline changed; and chapter 8 returns to this chapter's four gaps to state what the case can and cannot support.

Chapter reference list: `dissertation/references-ch2.md`. *Verification trail:* `dissertation/sources/verified.md` *(entries tagged* `[ch2]`*); works sought but not verified, and therefore not cited:* `dissertation/sources/unverified.md`.

3. Company and industry context

> **What this chapter establishes.** Which of the firm's assets actually transfer to a wellness-audio venture and which do not, and what kind of industry it would be entering — structurally unattractive on rivalry, buyers and substitutes, with a narrow defensible position built on rights plus measurement.

> **How to read the framework cells.** Every cell in the tables below carries the evidence it rests on. Verdicts (a VRIO "yes", a five-forces rating, a positioning coordinate) are the author's assessment *against that evidence*; where the repository carries no evidence capable of supporting a verdict, the cell reads *[to verify]* and the reader can see exactly which part of the analysis is unsupported. No cell is filled by inference from a competitor's marketing or from general industry knowledge.

3.1 Company profile

TDMusic Inc. (太簇音乐, Tempered Digital Music) is a technology-driven digital-content distribution and monetisation company headquartered in Beijing and operating across the United States, Europe, Japan and China. Company documents report revenue of $10.5 million, net income of $2.2 million, a $70 million valuation and 52 employees, an angel round led by Challenger Capital, and a market position described as top-three music distributor in China, Tier-1 YouTube partner and top supplier to Tencent Music (TDMusic Inc., 2022–2026). Its 2022 catalogue figures record approximately 106,000 songs in reserve and 75,000 released, approximately 85 per cent owned or directly licensed, 1.05 billion streams between January and July 2022, and distribution to more than 200 countries through over 220 digital service providers, including direct arrangements with Spotify, Apple and YouTube and endpoints such as Peloton, Tesla and Virgin Airlines.

Two further elements of the profile bear directly on the analysis that follows. The first is a peer-reviewed recommender: the firm's scenario- and preference-based auto-matching engine is published as Lu et al. (2021) in *PeerJ Computer Science*. The second is a demonstrated operating pattern: a Kindle Direct Publishing pilot that mined demand signals, produced content with AI assistance, published it and measured the result, reporting click-through rates two to three times benchmark, 4.2–4.5 star ratings and titles profitable within two to four months. The venture examined in this dissertation applies the same engine — demand signal, AI production, measured library — to the firm's core asset in a new vertical.

Vintage and reconciliation. The company file flags four figures for reconciliation before submission: one founding date (2020 for the Chinese entity versus March 2021 in the business-model deck), one region count (four regions versus five countries), the catalogue and stream figures (2022, and therefore stale), and the scope of the Kurt Hugo Schneider relationship (China agency operations versus distribution rights). Each company figure is reported here with its vintage, and none is treated as current or used inside an arithmetic result.

3.2 Asset inventory and VRIO

The asset inventory is the "ability" half of the research question. Table 3.1 runs the seven assets recorded in the company file through VRIO — is the asset **V**aluable for this venture, is it **R**are, is it costly to **I**mitate, and is the firm **O**rganised to capture the value — and states the competitive implication that follows.

Table 3.1. *VRIO analysis of TDMusic's asset inventory against the wellness-audio venture*

Asset	Evidence in the repository	V	R	I	O	Implication for the venture
AI distribution and matching algorithm (BAS / xDeepFM recommender)	Published as Lu et al. (2021), *PeerJ Computer Science*, 7, e716; company claims ~10× marketing ROI; scenario- and preference-based auto-matching of music to users.	Yes	*[to verify]*	*[to verify]*	Yes	The same engine can match calming music to stress contexts, and the peer-reviewed paper is citable credibility. But the method is published, and the repository holds no benchmark against competing recommenders, so rarity and inimitability cannot be asserted. **Competitive parity plus credibility**, not advantage.
AI production pipeline (covers, artwork, voice synthesis, video)	AI-assisted covers, audio artwork and intro content; video pipeline with four of six steps automated; talknet2 voice synthesis mature for English; UE5 digital humans. The business-model deck notes AI tools for artists "especially in meditation and lofi genres".	Yes	*[to verify]*	*[to verify]*	Yes	Low-cost production of calm arrangements and adaptive variants — the closest existing capability to what the product needs. The description is 2022-era and the company file flags it for refresh, so no claim is made about its current state of the art.
Rights and catalogue (~106k reserve / 75k released, ~85% owned or direct)	100k+ songs, majority direct-licensed; rights-management and royalty infrastructure; composition rights including cover and adaptation rights for part of the catalogue.	Yes	Yes	Yes	*[to verify]*	The raw material for real-music calm content, and the one genuinely non-imitable asset here: rights are legally exclusive. But *organisation to capture* turns entirely on whether adaptation and derivative rights exist per track — the top legal assumption A5, posterior 0.65, legal check pending. **Potentially sustained advantage, gated on A5.**
Distribution and platform relationships (220+ DSPs)	Direct deals with Spotify, Apple and YouTube; Peloton, Tesla and Virgin Airlines as distribution endpoints; top-three distributor in China, Tier-1 YouTube partner, top TME supplier (2022 figures).	Yes	Yes	*[to verify]*	Yes	Global launch capability on day one, and the Peloton-type wellness endpoints are directly relevant B2B2C precedents already in house. Inimitability is unclear: relationships are contractual and can be built with scale, and the repository holds no evidence on switching costs.
Demand-signal method (the KDP publishing pilot)	Reddit demand mining → AI production → publish → measure. Reported click-through 2–3× benchmark, 4.2–4.5★ ratings, titles	Yes	*[to verify]*	*[to verify]*	Yes	The continuity argument of the whole thesis: the same engine, a new vertical. As a repeatable process rather than a protected asset, its rarity and inimitability

Asset	Evidence in the repository	V	R	I	O	Implication for the venture
	profitable in two to four months.					are unevidenced — the repository records the pilot's results, not a comparison with anyone else's.
Artist network	Kurt Hugo Schneider (13.4M YouTube subscribers — relationship scope flagged for clarification), Cécile Corbel, regional artists across the US, EU and Asia.	Yes (for concept C3)	*[to verify]*	*[to verify]*	*[to verify]*	Artist-branded calm content is the credibility Endel lacks, and is the cheapest test of real-music demand. Every VRIO column except value is unverifiable because the company file itself flags the relationship scope (China agency operations versus distribution rights) as unresolved.
Cash engine and multi-region operations	$2.2M net income; 52 employees; teams in four regions; $10–20k ring-fenced for this phase.	Yes	No	No	Yes	Not an advantage — profitability is neither rare nor hard to imitate — but it is the *organisational condition* for honest exploration: Horizon 1 funds Horizon 2, and a ring-fenced budget makes a kill decision survivable.

Source. Assets and evidence transcribed from `company/COMPANY_BACKGROUND.md` §2; A5 posterior from `research/DECISION_MODEL.md` §3; budget from `DESIGN_THINKING_FRAMEWORK.md` framing update. VRIO verdicts: 28 cells (7 assets × V, R, I, O) — 17 supported by the cited evidence, 11 marked "to verify".

Reading of the VRIO. Only one asset — rights and catalogue — clears value, rarity and inimitability together, and its fourth column is exactly the question the pending legal check answers. That is a strategically uncomfortable result: the firm's single defensible asset is the one whose usability is unknown, and the user evidence in chapter 5 says that asset is the wrong content for the moment of highest need. Chapter 6 resolves this by segmenting the product rather than the asset.

3.3 Macro environment: PESTEL

Table 3.2. *PESTEL analysis of the consumer digital-wellness and functional-audio environment*

Factor	Evidence and reading	Source and caveat
Political / regulatory	Three regimes draw the same line differently. *US:* the FDA's two-factor general-wellness test permits "helps you relax" and "supports stress management" for low-risk lifestyle products, while "treats anxiety disorder" triggers software-as-a-medical-device status (U.S. Food and Drug Administration, n.d.). Medicare began reimbursing FDA-cleared digital mental-health devices on 1 January 2025 (HCPCS G0552–G0554), with seven apps qualifying, but clinicians must front the device cost (Healthcare Brew, 2025). *Germany:* 60 DiGAs listed, 31 of them mental health — the largest category (Bundesinstitut für Arzneimittel und Medizinprodukte, 2026). *UK:* NICE Early Value Assessment HTE9 provisionally recommends named anxiety apps pending evidence (National Institute for Health and Care Excellence, 2023). *China:* AI-driven diagnosis or treatment-recommendation software risks Class III classification; in one NMPA review only 5 of 61 surveyed products were classified at Class II (TMTPost, 2023).	`regulation-verified.md` rows 1, 6–7, 10; `SYNTHESIS` §4. A January 2026 FDA guidance update is described by law-firm secondary sources as broadening enforcement discretion; the primary text was not read and the update is treated as indicative only.

Factor	Evidence and reading	Source and caveat
Economic	Category size is contested rather than known: meditation apps $2.20 bn (2025) forecast to $6.99 bn by 2033, CAGR 14.67%, North America 43.22% share (Grand View Research, 2025); the broader mental-health-apps category $7.48 bn (2025) to $35.29 bn by 2034, CAGR 19.23% (Fortune Business Insights, 2025). These are different definitions and must not be summed. Against those forecasts, both leaders shrank in 2025: Calm approximately $210M, −24% year on year, approximately 3.5M subscribers (−500k); Headspace approximately $140M ARR against approximately $348M in 2024 (estimates conflict). Willingness to pay is market-dependent: Calm $69.99/year against Tide's ¥218/year (approximately $30, about 40%). German DiGA launch prices median €514 fall about 50% to approximately €221 after negotiation (Prova Health, 2025).	`market-wtp-verified.md` rows 1–2, 7; `competitors-verified.md` rows 1–2, 10, 29, 40; `SYNTHESIS` §3. Both market-size sources are modelled figures whose pages returned HTTP 403 on direct fetch and were corroborated via search.
Social	Prevalence: 359 million people with an anxiety disorder in 2021, 27.6% receiving treatment (World Health Organization, 2023); US 19.1% past-year and 31.1% lifetime on NCS-R fieldwork from 2001–2003 (National Institute of Mental Health, n.d.); China 5.0% 12-month and 7.6% lifetime (Huang et al., 2019) — roughly a fourfold gap most plausibly explained by under-diagnosis, stigma and instrument differences rather than by true prevalence. Behaviour: a March 2025 survey of 1,000 US adults found 92% said music helped them through hard times, 55% specifically for anxiety and 51% had substituted music for therapy (Tebra, 2025); an analysis of 110,000+ tracks on public "mental health" playlists found metal, not ambient, the most common genre. Engagement: approximately 4.7% 30-day retention, 1–4 lifetime sessions (Creswell & Goldberg, 2025).	`SYNTHESIS` §3, §6; `market-wtp-verified` rows 4–6; `user-sentiment-verified` via `SYNTHESIS` §6. China's ¥2.31 trillion "emotion economy" figure is real but bundles blind boxes, pets and escape rooms — it is not an app market.
Technological	The binding constraint is what wearables actually expose. Heart rate streams every 1–5 seconds during a session and is accurate (approximately 6% error; resting-HR MAPE 5.91%) (Hernando et al., 2018). HRV does not: HealthKit samples it opportunistically with lag up to approximately 30 minutes (Apple Inc., n.d.-a), and Apple Watch Series 9/Ultra 2 underestimated HRV by a mean 8.31 ms (MAPE 28.88%) against a Polar H10 reference, failing pre-set equivalence bounds (Apple Watch Series 9 and Ultra 2 heart-rate-variability validation study, 2024). Oura returns 5-minute HRV only from sleep, the following morning, with no streaming endpoint, and since 2024 partner apps get no data for Gen3+ users without an active membership (Oura Health, 2024). Platform-level substitution is real: Apple's free Vitals app ships a sleep score and overnight vitals with every Series 9+ watch (Apple Inc., n.d.-b).	`wearables-biofeedback-verified.md` rows on *Sensors* (2024), Hernando et al. (2018), Oura API v2 and membership gating; `competitors-verified` row 27; `ML_RESEARCH_DESIGN` §1–2.
Environmental	**Not material to this venture.** The product is software distributed through existing app stores and streaming infrastructure; it adds no manufactured hardware, no logistics and no physical packaging. The repository contains no environmental data, no carbon accounting and no environmental claim, so no environmental factor is asserted here in either direction — the cell is complete precisely because the honest answer is that this dimension does not bear on the decision.	Stated as "not material" per the dissertation brief; no repository source makes an environmental claim. Any future hardware partnership (cf. Calm × Ozlo earbuds) would reopen this cell.
Legal	*Rights.* Adaptation and derivative rights per catalogue track are unaudited; masters and publishing must be confirmed separately. This is assumption A5 (prior 0.50, posterior 0.65) and the legal check on ≥ 50 tracks carries the only decisive pending likelihood ratio (8.0 / 0.15). *Data.* GDPR lawful basis is consent, with EU data stored in the EU; HealthKit terms forbid advertising use of	`DECISION_MODEL` §3 A5 and §4; `ML_RESEARCH_DESIGN` §6; `regulation-verified` rows on the FTC rule and the Advertising Law; `SYNTHESIS` §4. No China enforcement precedent against a

Factor	Evidence and reading	Source and caveat
	health data; Oura's API is membership-gated. The FTC's Health Breach Notification Rule, broadened in April 2024, covers health and wellness apps regardless of FDA device status — avoiding SaMD status does not exempt a product from US health-specific regulation (Federal Trade Commission, 2024). *China.* Advertising Law Article 17 bars non-medical advertisers from referencing disease-treatment function, with penalties under Articles 58–59 of one to three times advertising cost or RMB 100,000–200,000 (Advertising Law of the People's Republic of China, 2015).	consumer anxiety wellness app was found — the risk is inferred from adjacent cases, not observed.

Source note. PESTEL: 6 of 6 cells sourced; 0 marked "to verify". The environmental cell is sourced as a reasoned "not material" rather than as a data point.

3.4 Industry structure: Porter's five forces

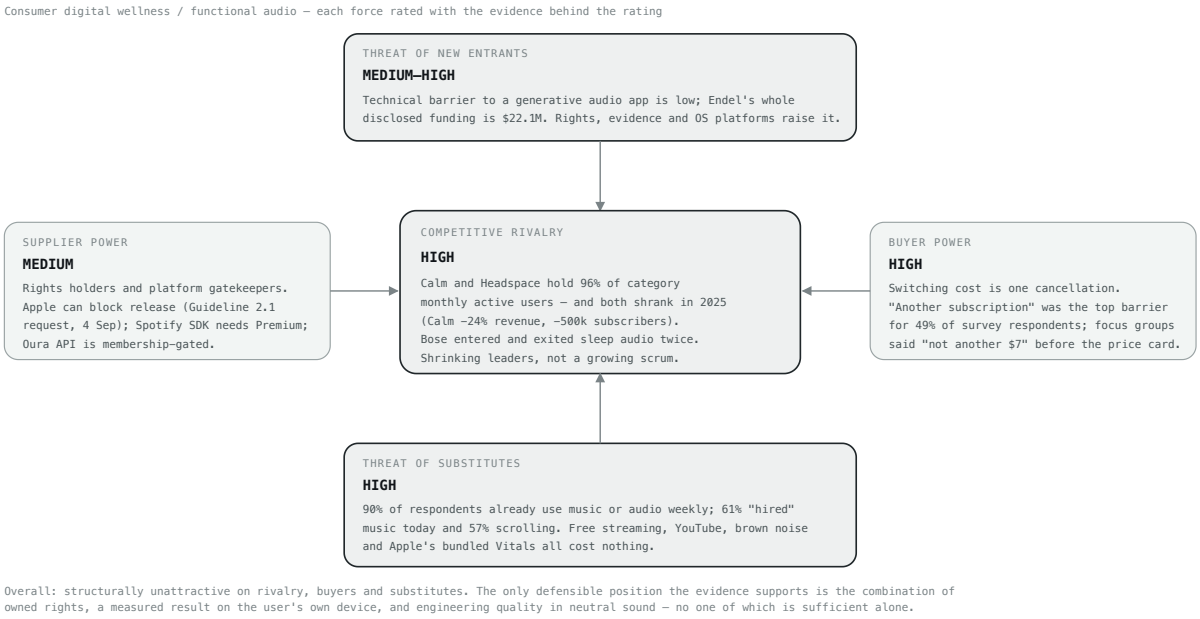


Figure 3.1. *Porter's five forces.* Ratings are the author's assessment against the cited evidence. Rivalry: `SYNTHESIS` §2, §6 and `competitors-verified` rows 1–2, 10, 37. Buyers: survey block E5 and focus-group theme T6. Substitutes: survey music-use and "hired today" items; `competitors-verified` row 27. New entrants: `competitors-verified` row 12. Suppliers: `TODO` §14–15, `app/RECOMMEND_BRIEF.md`, `wearables-verified` Oura row.

Table 3.3. *Five-forces ratings, evidence and consequence for the venture*

Force	Rating	One-line evidence	Consequence for this venture
Competitive rivalry	High	Calm and Headspace jointly hold 96% of category monthly active users, and both declined in 2025 (Creswell & Goldberg, 2025; Business of Apps, 2026a, 2026b).	Do not compete on catalogue breadth or brand spend; compete on the measured result.
Buyer power	High	"Another subscription" was the leading barrier for 49% of survey respondents; both focus groups preferred the feature bundled inside something already paid for.	Strengthens concept C2 (licence into a device or platform) relative to standalone B2C.

Force	Rating	One-line evidence	Consequence for this venture
Threat of substitutes	High	90% already use music or audio at least weekly; 61% "hired" music for calming today; Apple ships Vitals free with every Series 9+ watch (Apple Inc., n.d.-b).	The product must beat *free music the user already has*, not beat Calm.
Threat of new entrants	Medium–high	Endel's entire disclosed funding is $22.1M across five rounds (TechCrunch, 2022); the build barrier is low. Rights, an evidence base and distribution are the only real barriers.	Speed matters less than evidence; the pilot is a moat-building activity, not a compliance exercise.
Supplier power	Medium	Apple App Review issued a Guideline 2.1 information request on the first submission; Spotify playback requires Premium; Oura gates its API behind an active membership (Oura Health, 2024). Against that, approximately 85% of the catalogue is owned or directly licensed, internalising the music supplier.	Platform risk is real and current; music-supplier risk is largely internalised — if A5 clears.

Source note. Five forces: 5 of 5 rated with cited evidence; 0 "to verify".

3.5 Competitor analysis

3.5.1 Competitor profiles

Table 3.4. *Competitor profiles: ownership, model, wearable integration, independent evidence and documented weakness*

Player	Owner / funding	Model and pricing	Wearable integration	Independent evidence	Documented weakness
Calm	Private; $218M raised; $2bn valuation (2020 Series C), no newer disclosure (Tracxn, 2026)	$14.99/mo, $69.99/yr, $399.99 lifetime; B2B2C "Calm Health" claims 39M+ covered lives across 100,000+ organisations (Calm Health, 2026)	None disclosed; partnered with Ozlo earbuds (September 2025) (TechCrunch, 2025)	One RCT worldwide, no disclosed conflict of interest (O'Daffer et al., 2022)	2025 revenue approximately $210M, −24% year on year; approximately 3.5M subscribers, −500k; complaint patterns centre on billing, not content (Latka, 2026; Business of Apps, 2026a)
Headspace	Merged with Ginger (2021) into Headspace Health at a $3bn valuation; approximately $216M + approximately $220M prior funding (Healthcare Dive, 2021)	Consumer subscription plus B2B2C; approximately 2,700 employer and health-plan partners, approximately 100M lives. Consumer price point: *[to verify]*	None disclosed	14 RCTs — the largest base in the category — but 7 of 14 disclose a conflict of interest and only 36% were pre-registered (O'Daffer et al., 2022)	Estimated revenue fell from approximately $348M (2024) to approximately $140M ARR (2025); 13% workforce cut November 2024; estimates conflict sharply (Behavioral Health Business, 2024; Business of Apps, 2026b)
Endel	Private; approximately $22.1M total,	Subscription. Valuation approximately	Real-time Apple Watch heart rate (not HRV) plus	Sole peer-reviewed study is Endel- and Arctop-funded with	Its own chief commercial officer concedes

Player	Owner / funding	Model and pricing	Wearable integration	Independent evidence	Documented weakness
	$15M Series B (April 2022, Waverley Capital and True Ventures) (TechCrunch, 2022)	$47.5M and revenue approximately $15.8M trace to a single uncited aggregator page — low confidence. Price point: *[to verify]*	motion, time and light; Apple Watch App of the Year 2020. Oura integration announced but not independently confirmed	Arctop-employed authors and measures focus, not anxiety (Haruvi et al., 2022)	algorithmic audio "is paid the same" as composed music despite being "very different"; the artist deals are revealed evidence that generative sound under-satisfies (Blakeley, 2024)
Brain.fm	Private; disclosed funding implausibly small (approximately $125k seed plus an NSF grant of unconfirmed size) — likely incomplete records	Approximately $9.99/mo or approximately $99.99/yr; third-party sources show inconsistent historical pricing ($6.99/mo, $49.99/yr elsewhere)	None found	Peer-reviewed *Communications Biology* study with Northeastern's MIND Lab on attention-network engagement (Woods et al., 2024)	Proprietary stimuli and a commercial conflict of interest inside its own validation study; the study measures attention, not anxiety
Tide (潮汐)	China; "near ¥10M" Pre-A led by Panda Capital; approximately 30-person team (VBData, 2026)	Subscription-only, no advertising: ¥218/yr (approximately $30) or ¥27/mo, plus a ¥8/mo sound-card tier	None found	None found	Structurally sub-scale; growth claims are founder-reported and unaudited; the domestic sleep-app sector is described by trade press as having gone "from hot commodity to dispensable" (36Kr, 2026)
MedRhythms	Private; $34M total funding; CEO replaced July 2025 by a former public-medtech CEO	Prescription digital therapeutic; Medicare DME code HCPCS E3200, final payment determination effective 1 April 2025 (PR Newswire, 2025). List price: *[to verify]*	Proprietary sensor-embedded device, not consumer wearables	The only player here with an FDA regulatory pathway: Breakthrough Device Designation (June 2020) for InTandem (PR Newswire, 2020); MOVIVE holds a lower Class II Rx-only listing	Narrow indication (stroke and Parkinson's gait, not anxiety); five years from designation to reimbursement

Source. `competitors-verified.md` rows 1–24, 28–29; `SYNTHESIS` §2; `market-wtp-verified` rows 7, 49–51. Every dollar figure for Calm, Headspace and Endel traces to third-party aggregators (Latka, Business of Apps, Tracxn) that disagree with each other by wide margins; the *direction* is consistent, the precision is not.

3.5.2 Competitor marketing mix (4P)

Table 3.5. *Competitor marketing mix*

Player	Product	Price	Place	Promotion
Calm	Meditation, Sleep Stories, music; standalone Calm Sleep app (September 2025) with 300+ hours of sleep content and 500 Sleep Stories; Calm Health clinical programmes rooted in CBT, DBT and ACT with PHQ/GAD screening	$14.99/mo · $69.99/yr · $399.99 lifetime	iOS and Android stores; employer and health-plan channels (Solera network, January 2026); Ozlo earbuds hardware bundle	Hardware bundle discount ($50 with Ozlo) and press. Channel mix and spend: *[to verify]* — no marketing data in the repository
Headspace	Meditation plus Ginger coaching and therapy inside Headspace Health	*[to verify]* — the repository records the B2B2C model and subscriber counts but no consumer price	Approximately 2,700 employer and health-plan contracts covering approximately 100M lives; consumer app stores	*[to verify]*
Endel	Generative soundscapes by mode; artist collaborations (Grimes, James Blake, Miguel); UMG partnership (May 2023) using catalogue stems; Warner's Spinnin' Records committed to 50 AI-generated wellness albums	Subscription; price point *[to verify]*	iOS, Android, macOS and Apple Watch (standalone and companion modes)	Label and artist partnerships *are* the promotional engine: the UMG and WMG deals generate the coverage (Music Business Worldwide, 2023; PR Newswire, 2023)
Brain.fm	Patented amplitude-modulated functional music for focus, relaxation and sleep	Approximately $9.99/mo · approximately $99.99/yr (fluctuates)	App stores and web; direct vendor pricing page	The peer-reviewed study is the marketing asset — announced via press release framed as "world's first science-backed purpose-built focus music"
Tide (潮汐)	White noise, meditation, sleep and focus content; subscription-only with explicitly no advertising	¥218/yr · ¥27/mo · ¥8/mo sound card	Chinese app stores; partnerships with Apple (employee-wellness procurement), Santonban coffee, Atour Hotels and Lululemon	Brand partnerships and founder interviews; spend and channel mix *[to verify]*
MedRhythms	InTandem (MR-001) rhythmic auditory stimulation for chronic-stroke gait; MOVIVE (MR-005) for Parkinson's gait	Medicare-reimbursed under HCPCS E3200 from 1 April 2025; list price *[to verify]*	Clinician prescription; national distribution agreement with Edwards Health Care Services (September 2025)	Regulatory milestones and the UMG catalogue partnership; no consumer promotion

Source note. Competitor 4P: 24 cells (6 players × 4 Ps) — 18 sourced, 6 marked "to verify". Competitor profiles: 36 cells, 33 sourced, 3 marked "to verify". Sources as §3.5.1 plus `competitors-verified` rows 5–7, 14, 19, 23, 29.

3.5.3 Positioning map

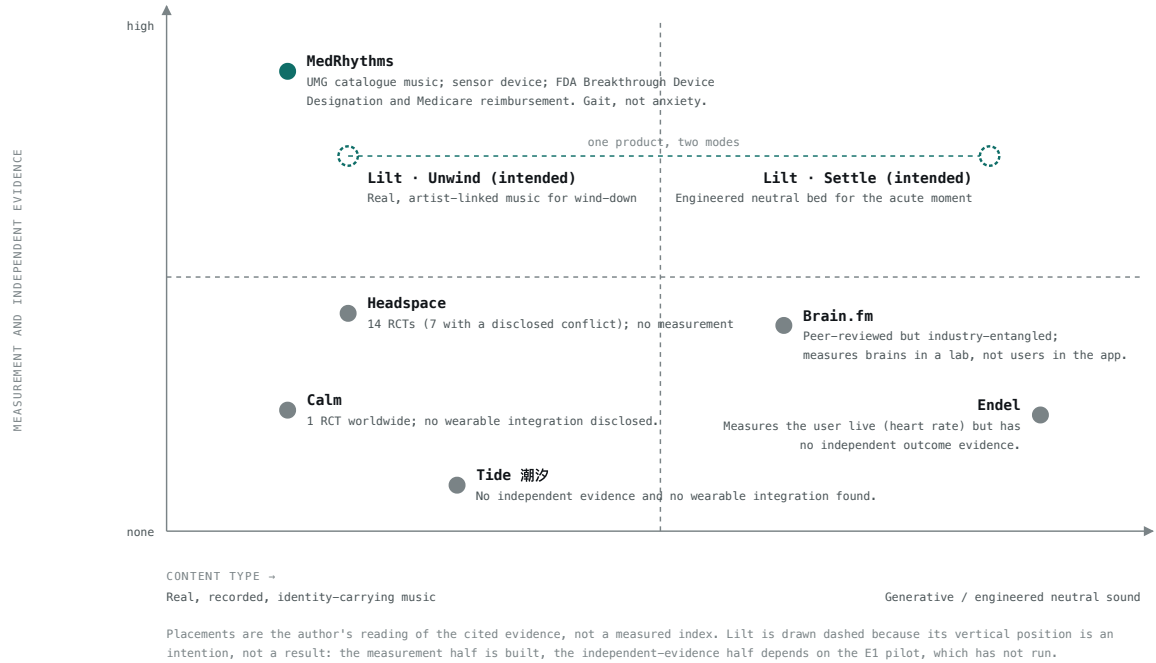


Figure 3.2. *Positioning map.* Horizontal axis: content type, from real recorded music to generative or engineered neutral sound. Vertical axis: whether the product measures the user's physiology and whether independent peer-reviewed outcome evidence exists. Sources per competitor as §3.5.1. Lilt's two modes are drawn as one dashed span because the strategic proposition is precisely that a single product should occupy both ends of the content axis, segmented by arousal state (chapter 5).

The map makes the strategic opening legible. The upper half is nearly empty: only MedRhythms combines real music with a regulatory-grade evidence position, and it does so in a narrow, prescription-gated, non-anxiety indication. Endel occupies the measurement corner without independent evidence; Brain.fm has evidence without user measurement; Calm, Headspace and Tide have neither. The position this project intends — measured on the user's own device, evidenced by a pre-registered trial, and content-segmented by arousal state — is unoccupied. It is also unproven, which is why the marker is dashed.

3.6 Market sizing: TAM, SAM and SOM

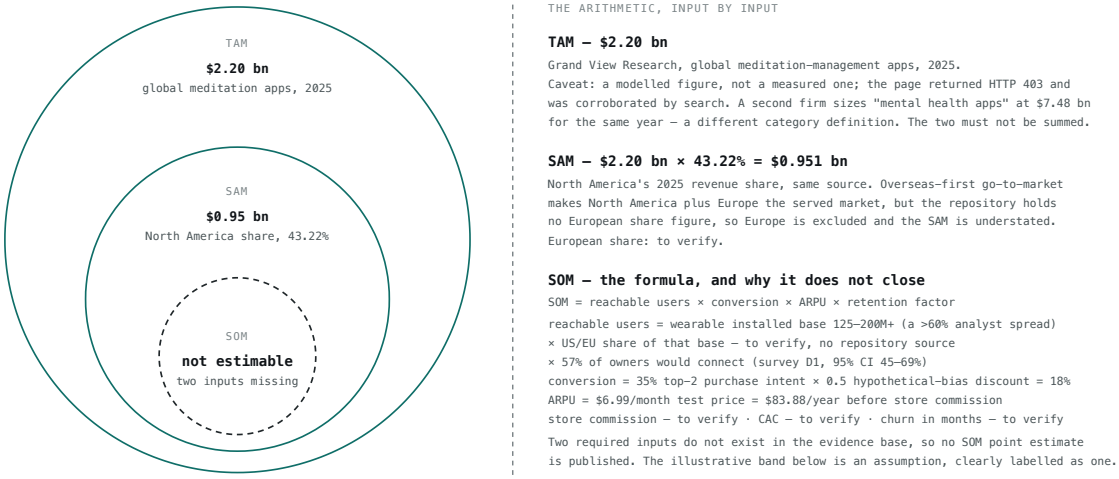


Figure 3.3. *TAM / SAM / SOM with the arithmetic shown.* Every input is labelled with its source and its caveat. The SOM circle is drawn dashed and set apart because the project declines to publish a point estimate it cannot support: two of the formula's inputs (the US/EU share of the wearable installed base, and customer acquisition cost) have no repository source at all.

```
TAM = $2.20 bn   global meditation-management apps, 2025
      [Grand View Research; a modelled figure; page 403 on fetch, corroborated by
search]
      ≠ $7.48 bn "mental health apps", same year, a wider definition
      [Fortune Business Insights — a different market. The two must not be summed.]

SAM = $2.20 bn × 0.4322 = $0.951 bn   North America revenue share 43.22%
      [same source; Europe excluded — the repository gives no European share, so SAM
is understated]

SOM = reachable × conversion × ARPU × retention
      reachable  = 125–200M+ Apple Watch installed base
                   [analyst estimate, > 60% spread, not company-disclosed]
                 × US/EU share of that base   to verify — no source
                 × 0.57 would connect   [survey D1, owners, 95% CI 0.45–0.69]
      conversion = 0.35 top-2 intent × 0.5 discount = 0.18
                   [survey F8 at $6.99; Morwitz et al. hypothetical-bias rule]
      ARPU       = $6.99/mo = $83.88/yr before commission
                   [test price; PSM range $5.19–$8.46, OPP $6.28, IPP $7.37]
      retention  = 4.7% D30 base rate; 15% target   [Creswell & Goldberg, 2025; A6 =
0.25]
      CAC        = to verify — the landing A/B has a budget and no result yet

⇒ SOM is not computed.  Two of the inputs have no source at all.
   Illustration only, not a finding: 1% of the North American meditation-app
   market = $9.5M annual revenue. The 1% is an assumption chosen for illustration.
```

Source. Inputs: `market-wtp-verified` rows 1–2, 49, 51, 57–58; `SYNTHESIS` §3, §6; `surveys/analysis/results.json` via `survey-results.html` (`wearable.d1_top2_owners`, `intent`, `vw.US_EU`); wearable installed-base estimate from Above Avalon (2025). TAM/SAM/SOM: 9 inputs sourced, 5 marked "to verify" (European revenue share, US/EU share of the wearable installed base, store commission, CAC, churn in months).

> **Why the "emotion economy" numbers are not used.** China's headline figures — RMB 2.31 trillion "emotion economy" in 2024, forecast above RMB 4.5 trillion by 2029, and RMB 495.58 bn "sleep economy" in 2023 — are real, verified iiMedia figures (iiMedia Research, 2023, 2025), but they bundle mattresses, bedding, blind-box toys, pets and escape rooms. They are not app-market sizes and are excluded from every calculation above. They appear in this dissertation only as context for the observation that Chinese willingness to pay for wellness *content* is structurally low.

3.7 SWOT and the strategic issue statement

Table 3.6. *SWOT analysis*

	Helpful to the objective	**Harmful to the objective**
Internal	**Strengths.** *Owned rights at scale* — approximately 106k songs in reserve, approximately 85% owned or directly licensed, with royalty infrastructure. *Day-one global distribution* — 220+ DSPs including Peloton, Tesla and Virgin Airlines endpoints. *A peer-reviewed recommender* — Lu et al. (2021). *A proven engine* — the KDP pilot: click-through 2–3× benchmark, profitable in 2–4 months. *Cash to fund exploration* — $2.2M net income; $10–20k ring-fenced for this phase. *A working product already* — v0.5.1 on the web, TestFlight builds, 1.0 submitted to the App Store with nine owned tracks.	**Weaknesses.** *No efficacy evidence of its own* — E1 has not run; A2a 0.85 and A2b 0.56 are borrowed from meta-analyses. *The core asset is the wrong content for the core moment* — P(A1a) ≈ 0.08. *Rights for adaptation unaudited* — A5 0.65; per-track derivative rights flagged unknown in the company's own file. *Company figures stale and unreconciled* — founding date, catalogue and stream counts (2022), artist-relationship scope. *Insider-researcher position* — the founder is also the analyst; bias controls are declared but the conflict is structural. *A known product inconsistency* — the iOS build ships no Bluetooth plugin while the store description mentions a chest strap; to be corrected in 1.1. *HRV cannot be steered live* on any consumer API, so the "closed loop" is heart-rate-driven by necessity.
External	**Opportunities.** *The action gap* — 38% of survey respondents do nothing after a high stress reading; focus-group owners read the score "like the weather". This is the hole a closed loop fills. *A concrete unmet spec* — complaints about existing neutral audio are specific and engineerable: sharp frequencies, low hum causing nausea, "fake healing" reactance, beat entrainment. *A real wind-down segment for real music* — 47% [39–56%] chose the real song for wind-down; 35% rated real artists top-2 in value. *An evidence vacuum in the category* — no independent review exists for Endel, Brain.fm or any Chinese competitor. *B2B2C routes already in house* — Peloton-type endpoints, plus documented label appetite for functional music (UMG and WMG with Endel).	**Threats.** *The retention cliff* — approximately 4.7% D30 industry-wide, 1–4 lifetime sessions; A6 posterior 0.25 (Creswell & Goldberg, 2025). *A maturing, shrinking category* — Calm −24% revenue and −500k subscribers in 2025; Headspace revenue down sharply with layoffs. *Platform disintermediation* — Apple ships Vitals free with every Series 9+ watch; Oura gates its API behind a membership. *Regulatory claim creep* — any "treat anxiety" wording triggers SaMD status in the US and Advertising Law Article 17 exposure in China. *The DTx commercial precedent* — Pear filed Chapter 11 in April 2023 (Fierce Biotech, 2023) and Akili was taken private at $0.434/share in mid-2024 (Virtual Therapeutics, 2024), both on reimbursement rather than efficacy. *Gatekeeper risk is live* — the first App Store submission drew a Guideline 2.1 information request. *Subscription fatigue* — 49% named "another subscription" as their barrier.

Source note. SWOT: 25 cells, all 25 sourced, 0 "to verify". Sources: `COMPANY_BACKGROUND` §1–2, §4; `DECISION_MODEL` §3; `SYNTHESIS` §2–§4, §6; survey results as published; focus-groups `SYNTHESIS` T3, T6–T7; `TODO` §11, §14–16; `wearables-verified`; `regulation-verified`.

Strategic issue statement. Bringing the four analyses together yields one issue, and it is not the issue a conventional attractiveness screen would return. The industry is structurally unattractive on three of five forces, the two category leaders are shrinking, and no serviceable obtainable market can be computed from the available inputs; on those grounds alone the venture would not be

recommended. What makes it worth examining is the intersection of an empty quadrant in Figure 3.2 with an asset base that only partly reaches it: the firm holds exactly one asset that clears value, rarity and inimitability together — rights and catalogue — and that asset's usability is gated on an unaudited legal question (A5) while the user evidence in chapter 5 indicates it is the wrong content for the moment of highest need. The strategic issue for TDMusic is therefore not "should we enter consumer wellness audio" but: *under what evidentiary conditions, and in which of several concept forms, can a rights-holding distributor convert an asset that is defensible but possibly mis-matched into a defensible position in a category where the only unoccupied ground is measured, evidenced adaptation?* Chapters 5 and 6 answer the first half; chapter 9 states the conditions.

4. Methodology

> **What this chapter establishes.** That the study is insider action research conducted under a pragmatist commitment; that its evidence base is a convergent mixed-methods design in which every source has one declared job and none is counted twice; that the decision rule — priors, likelihood ratios, quality shrinkage and gate thresholds — was written down before any of the data arrived; and that the identification strategy for the efficacy claim is a within-subject crossover whose power limits were declared in advance rather than discovered afterwards.

4.1 Research philosophy and the insider's position

The philosophical commitment is pragmatist: the study is organised around what would have to be true for a decision to be justified, and it selects methods by their fitness for answering that question rather than by allegiance to a single paradigm. Pragmatism is the position that permits a genuinely mixed design in which qualitative and quantitative evidence are combined without either being reduced to the other (Creswell & Plano Clark, 2018). It is also the position that makes the dissertation's central methodological device — an explicit, published probability attached to each contestable belief — a reasonable thing to do rather than an affectation: the number exists so that the decision it supports can be argued with.

The frame is insider action research (Coghlan, 2019). The author is the founder of the firm and simultaneously the researcher; he diagnoses, plans, acts, evaluates and learns inside the organisation he owns. This is a methodological commitment rather than a label, because it changes what rigour consists of. An outsider's rigour is distance: the researcher has no stake in the result and the reader's confidence rests on that. An insider has a stake that cannot be removed, so his rigour must be *declared procedure* — thresholds written before data, kill criteria written down, evidence weights fixed by a published rubric, and every pre-registered test reported whether it fires or not. Reflexivity and bias control are therefore part of the method and are set out in §4.6 rather than being confined to a limitations paragraph, and the author's influence on the work is examined again, as a substantive matter, in §8.4.

The insider position is also, in this case, the reason the study is possible at all. The evidence reported here includes the firm's asset inventory, its rights position, its internal comparison of per-song revenue, and a working product built inside a three-month window; none of that is available to an outside researcher. The methodological question is not whether to accept the insider's access but how to bound the insider's judgement, and that is what the decision layer in §4.4 is for.

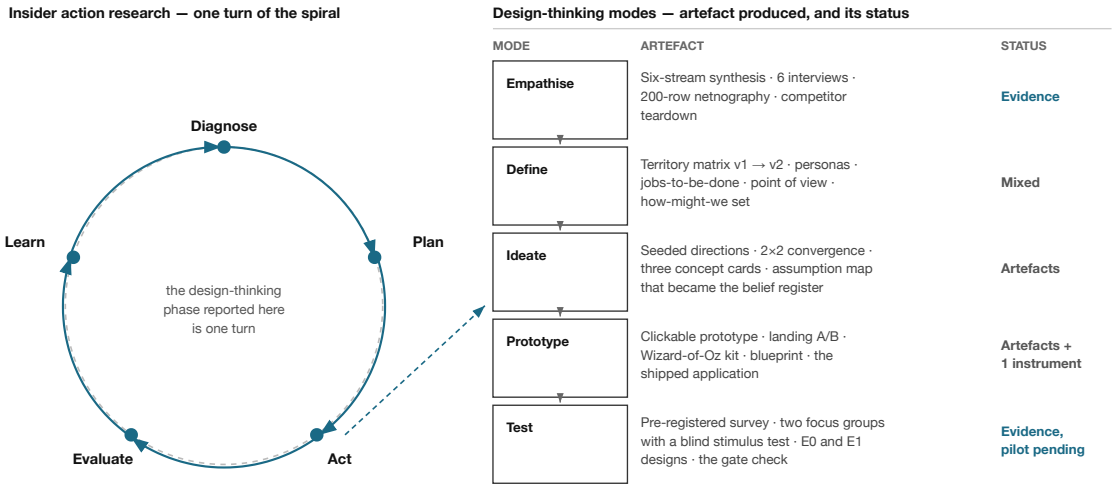
Three positional facts are stated because they shape what follows. The author has a commercial interest in a PROCEED outcome; the assets whose fitness is being tested are his firm's; and the same person specified the belief register, chose the priors and wrote the gate thresholds. Against these, the controls listed in §4.6 are implemented rather than aspirational — the research assistant scripts, the analyst blinding, the pre-registration, the reporting rule — and the residual conflict is conceded in §8.4 as structural rather than removable.

4.2 Research design: design thinking as the action cycle, mixed methods as the evidence base

4.2.1 The action-research cycle and the design-thinking modes

Action research proceeds as a spiral of diagnose → plan → act → evaluate → learn. Design thinking supplies the pacing of one turn of that spiral: the d.school's five modes — empathise, define, ideate, prototype, test — sequenced by the Double Diamond's alternation of divergence and convergence. The contribution of the design-thinking apparatus is not creativity but discipline: it forbids convergence before the problem space has been genuinely opened, and it requires that each mode terminate in a named artefact (Brown, 2008; Liedtka, 2015).

Figure 4.1 The action-research cycle and the design-thinking modes



The status column is what keeps an artefact from being read as a finding: personas and concept cards are outputs of the process, not results of it, and the chapter that reports findings draws only on the rows marked Evidence.

Sources: DESIGN_THINKING_FRAMEWORK.md §1–5; research/METHODOLOGY_DESIGN.md §1.

Figure 4.1. *The action-research cycle mapped onto the design-thinking modes, with the artefact each mode produced and the evidence status of that artefact.* The outer ring is the action-research spiral; the inner sequence is one turn of it, run as a design-thinking phase. The right-hand column records the epistemic status of each output, which is what prevents an artefact from being read as a finding. Sources: `DESIGN_THINKING_FRAMEWORK.md` §1–5; `research/METHODOLOGY_DESIGN.md` §1.

The distinction between *artefact* and *evidence* is enforced throughout and is worth stating plainly, because design-thinking write-ups frequently blur it.

Table 4.1. *The design-thinking modes, the artefacts each produced, and the epistemic status of each*

Mode	Artefacts produced	Epistemic status
Empathise	Six-stream literature and market synthesis; six interviews; a 200-row netnography; a ten-product competitor teardown	**Evidence.** The synthesis follows the project's fact-checked verification files, using corrected figures and flagging unsupported claims
Define	Territory scoring matrix v1 → v2 (Monte-Carlo re-score); personas; jobs-to-be-done;	**Mixed.** The matrix is evidence-driven and was re-scored after the interviews. The personas, JTBD, POV and HMW statements are

Mode	Artefacts produced	Epistemic status
	point-of-view and how-might-we statements	written in the framework document as illustrative examples rather than as coded outputs, and are reported as artefacts only
Ideate	Ten seeded directions plus session output; a 2×2 convergence; three concept cards (C1–C3); an assumption map that became the belief register	**Artefacts.** The framework asks for "≥ 15 raw ideas"; no record exists of how many were actually generated, so no count is claimed as a result. The number of participants in the ideation session is **[to confirm with the author]**
Prototype	Clickable prototype; landing A/B with two headline variants; Wizard-of-Oz kit; product blueprint; the Lilt progressive web application and iOS build	**Artefacts**, plus one behavioural instrument (the landing page) whose result did not exist at the time of writing
Test	Survey v2 (seven blocks, pre-registered); two focus groups with a blind stimulus test; E0 calibration and E1 efficacy designs; the gate check	**Evidence**, with the efficacy pilot outstanding

4.2.2 Mixed-methods architecture

The evidence architecture is a convergent mixed-methods design with a sequential quantitative core (Creswell & Plano Clark, 2018). The qualitative strand runs interviews (n = 6) → netnography (n = 200 coded rows) → focus groups (2 groups × 6–8 participants). The quantitative strand runs literature priors → survey (n ≥ 100) → efficacy pilot (n ≥ 20). Behavioural evidence (a landing-page A/B) and a legal check on adaptation rights enter alongside both strands.

The two strands have different jobs. The qualitative strand *generates* hypotheses — the arousal-state split that reorganised the entire project came out of three interviews — *specifies* instruments (the nine-item acoustic-tolerance matrix in the survey is a transcription of interview complaints), and *interprets* quantitative results. The quantitative strand *quantifies* prevalence, shares and price points, and *tests* causal claims. Integration is not left to narrative: each source enters the decision layer as a likelihood ratio whose magnitude is bounded by that source's quality (§4.4), so nothing is double-counted and nothing is weighted by rhetorical force.

4.2.3 The triangulation matrix

The design principle is that **every source is assigned a role**, and only sources with the appropriate role may move a given belief. The matrix below is the operational statement of that principle. Roles are: *prior*, which sets a starting belief; *generate*, which produces hypotheses; *quantify*, which estimates a share or a mean; *test*, which estimates a causal effect; and *texture* or *probe*, which informs interpretation and enters only with heavy shrinkage.

Table 4.2. *The triangulation matrix: the role assigned to each source for each construct*

Construct	Literature	Interviews	Netnography	Focus groups	Survey	Pilot	Landing / legal
Need prevalence and intensity	prior	texture	texture	segment texture	**quantify**	—	—
Timing, place, triggers	—	generate	—	**map**	quantify	—	—

Construct	Literature	Interviews	Netnography	Focus groups	Survey	Pilot	Landing / legal
Under-served outcomes	—	generate	complaints	rank	**quantify** (ODI)	—	—
Content by state (A1a / A1b)	weak prior	**generate** (disconfirming)	both coexist	**stimulus test**	**quantify** (C3–C5)	**A/B test**	—
Efficacy (A2a / A2b)	**prior** (meta-analyses)	anecdote	anecdote	—	—	**test**	—
Engagement (A3)	precedent	absence	action gap	probe	quantify	think-aloud	**conversion**
Willingness to pay (A4)	market data	—	billing complaints	reaction	**quantify** (PSM, intent)	—	**conversion**, partner talks
Rights (A5)	label deals	—	—	—	—	—	**legal check**
Retention (A6)	base rate	—	quotes	—	—	—	(post-phase)

Transcribed cell for cell from `research/METHODOLOGY_DESIGN.md` *§1.1.*

The matrix earns its place by making double-counting visible: a construct whose row contains only "texture" cells has no source entitled to move a gate. Efficacy (A2) is the clearest case — literature may set its prior and only the pilot may test it, which is why the gate thresholds in §4.4 were deliberately placed beyond the reach of the literature.

4.3 Instruments and sampling

4.3.1 Instruments

Validated instruments are used verbatim wherever one exists for the construct; where none exists, the instrument is ad hoc, is declared as such, and is pre-registered before fielding.

Table 4.3. *Instruments by construct, with type and validity handling*

Construct	Instrument	Type	Validity handling
Anxiety symptoms	GAD-2 (Kroenke et al., 2007)	Validated	Verbatim; sensitivity and specificity known at cutoff ≥ 3
Perceived stress	PSS-4 (Cohen & Williamson, 1988)	Validated	Verbatim; α reported; convergent correlation with GAD-2 expected .5–.7
Current alternatives	Jobs-to-be-done "what is hired today" (Christensen et al., 2016)	Adapted	The competitive set is behaviours, not applications
Under-served outcomes	Outcome-driven innovation, importance × satisfaction (Ulwick, 2002)	Adapted	Outcome statements drawn from interview JTBD; opportunity = I + max(I − S, 0); threshold ≥ 6
Content preference by state	Within-subject scenario forced choice, two scenarios	Ad hoc, pre-registered	Scenario order randomised; McNemar (1947), exact test for small discordant counts; segmented by GAD-2
Acoustic tolerances	Nine-item −2…+2 matrix	Ad hoc	Items are interview complaints verbatim; used as an engineering specification, not as a scale

Construct	Instrument	Type	Validity handling
Acceptance	TAM perceived usefulness, ease of use, behavioural intention (Davis, 1989; Venkatesh & Davis, 2000)	Adapted, two items each	$\alpha \geq .70$ target; OLS with standardised betas; PLS-SEM as robustness if $n \geq 150$
Relative advantage	Diffusion-of-innovations relative advantage (Rogers, 2003)	Adapted, single item	Top-2-box share
Feature value	Kano functional / dysfunctional pairs (Kano et al., 1984) with Better/Worse coefficients (Berger et al., 1993)	Standard	Evaluation table; dominant category; category reported by segment
Price sensitivity	Price sensitivity meter (van Westendorp, 1976)	Standard	Cumulative curves; optimal price point, indifference price point, acceptable range; per currency; monotonicity enforced per respondent
Purchase intent	Five-point scale at a shown price	Standard	Top-2 box × 0.5 hypothetical-bias discount (Morwitz et al., 2007)
Behavioural WTP proxy	Landing-page visitor → waitlist conversion	Behavioural	≥ 5% pre-registered, per positioning variant
State anxiety (pilot primary)	STAI-S (Spielberger, 1983); six-item short form (Marteau & Bekker, 1992)	Validated	Pre and post each session; the primary outcome
Physiological state (pilot)	Heart rate (bpm) and RMSSD / SDNN (ms) from the wearable; Polar H10 sub-sample	Objective, noisy	Artefact rules, calibration, residualisation (§4.5)
Expectancy (pilot covariate)	Two items, including a credibility item (Devilly & Borkovec, 2000)	Adapted	Entered as a covariate to control expectancy
Qualitative themes	Reflexive thematic analysis (Braun & Clarke, 2006); focus groups (Krueger & Casey, 2015; Morgan, 1997); netnography (Kozinets, 2020); critical incidents (Flanagan, 1954)	Qualitative	Hybrid codebook; second coder on 20% of transcripts, Cohen's $\kappa \geq .70$; saturation when a group adds no new code

From `research/METHODOLOGY_DESIGN.md` *§2 and* `surveys/questionnaire-v2.md` *(analysis plan).*

The survey instrument is organised in seven blocks under a 5W1H frame — who (A); why, when and where (B); what works (C); how to prove it (D); why use and why not (E); how much (F); and voice (G) — and every block states, in the instrument itself, the construct it measures, the assumption it feeds, the statistic to be computed and the threshold that will trigger a likelihood ratio. That is what makes the survey a pre-registered instrument rather than a questionnaire: the decision consequence of each possible answer was fixed before fielding.

The focus groups are the one instrument designed specifically to *disconfirm*. Each group begins with a blind audio stimulus test — three sixty-second, loudness-matched clips presented in rotated order (S-A, a calm instrumental arrangement of a known song; S-B, an engineered neutral bed with no pitch centre, no rhythm and no content below 60 Hz; S-C, a guided voice over a soft pad) — rated privately on a card before any discussion, so that group dynamics cannot contaminate the individual judgement that matters most to the central hypothesis. Only afterwards is the content of each clip revealed and discussed.

4.3.2 Sampling and power

Survey. Target n ≥ 100, preferably 150, giving a margin of error of ±9.8 percentage points (n = 100) or ±8 points (n = 150) on a 50 per cent share. Segment cells are targeted at ≥ 40 wearable owners and ≥ 30 GAD-2-positive respondents; any cell below 30 is reported as directional. Channel of origin is tracked across LinkedIn, Reddit, WeChat and an EMBA cohort snowball, and post-stratification weights are applied only if a channel skew above 20 points is observed, with both weighted and unweighted results reported. Fielding stops at n ≥ 150 or at the deadline, whichever comes first, and explicitly never *because* the numbers look good — a pre-commitment against optional stopping.

Focus groups. Two to three groups of six to eight (Krueger & Casey, 2015): one of United States and European wearable owners online, one of Chinese office and hybrid workers offline, and an optional third for the sleep-onset segment. Their purpose is saturation and stimulus reaction, not estimation; recruitment stops when a group yields no new code.

Efficacy pilot. Within-subject, three sessions per participant. For a paired design at α = .05 two-sided and 80 per cent power, d_z = 0.5 requires n = 34; 0.6 requires 24; 0.65 requires 21; and 0.8 requires 15. The Cochrane perioperative estimate (−5.7 STAI-S units, SD ≈ 10, so d ≈ 0.55) implies n ≈ 28 for self-report, while physiological effects at d ≈ 0.4 require n ≈ 50. **Therefore n = 20 is a feasibility signal for self-report and is under-powered for physiology, and this dissertation says so before the data rather than after.** Two remedies are built into the design: three sessions per participant, which extracts more within-person information from the same recruitment, and a Bayesian re-analysis using the meta-analytic prior, in which the posterior rather than a p-value drives the gate.

4.4 The decision layer

4.4.1 Why beliefs, and the update rule

The reason for expressing the project's contestable claims as probabilities rather than as forecasts is Knight's (1921) distinction between risk, where the distribution of outcomes is known, and uncertainty, where it is not. An exploration of this kind sits at a level of uncertainty at which no defensible frequency exists for most of what matters, so the honest object of estimation is the decision-maker's own degree of belief, together with an explicit rule for how evidence should revise it (Courtney et al., 1997). Making that belief public and arithmetic is what allows a reader to disagree with it precisely rather than generally.

For an assumption H with prior probability P(H): prior odds are $O_o = P(H) / (1 − P(H))$; each evidence item E_i contributes a likelihood ratio $LR_i = P(E_i \mid H) / P(E_i \mid \neg H)$; posterior odds are $O = O_o \times \prod LR_i$; and the posterior probability is $O / (1 + O)$. A likelihood ratio above 1 supports the hypothesis, below 1 disconfirms it, and exactly 1 is uninformative. Independence between evidence items is assumed for the multiplication; where items are clearly correlated — three interviewees recruited through one channel, three meta-analyses that overlap in their primary studies — the likelihood ratios are shrunk toward 1 before multiplying rather than treated as independent.

The precedent for using subjective likelihood ratios to integrate qualitative and quantitative evidence in social-science research is explicit: Fairfield and Charman (2017) on explicit Bayesian analysis for process tracing, and Humphreys and Jacobs (2015) on a Bayesian approach to mixing methods. The evidence-quality grading is modelled on GRADE (Guyatt et al., 2008). The precedent for pre-registering kill criteria in an innovation decision is the lean-startup and business-experimentation literature (Ries, 2011; Bland & Osterwalder, 2019).

Figure 4.2 How one piece of evidence becomes a movement in a belief

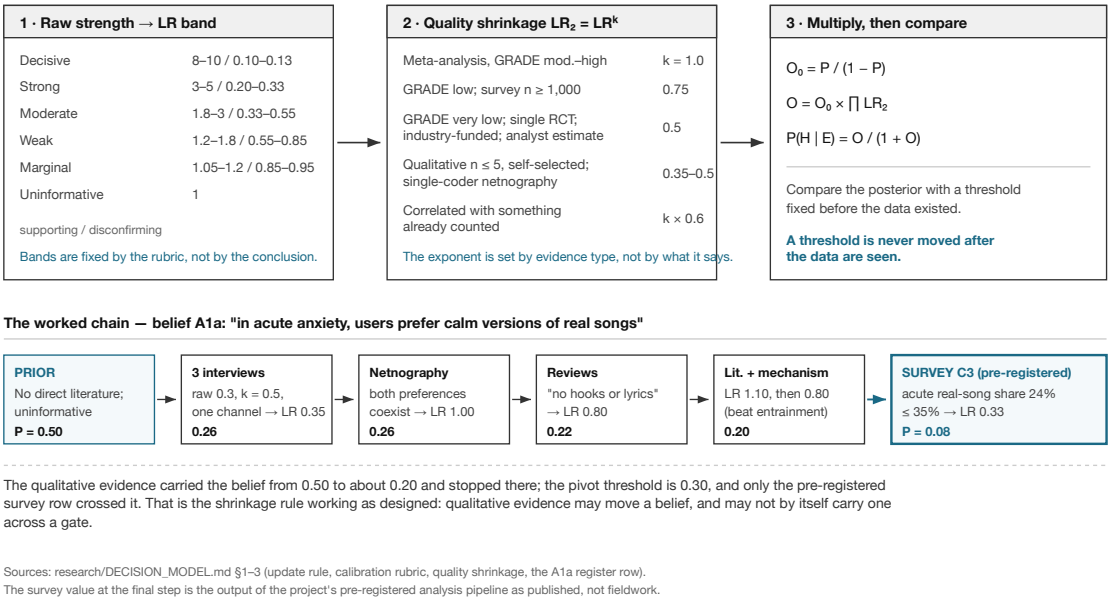


Figure 4.2. *How one piece of evidence becomes a movement in a belief.* Left: the raw strength of the evidence is mapped onto a likelihood-ratio band by the calibration rubric. Centre: the band is shrunk by an exponent set by the evidence type, and again for correlation with anything already counted. Right: the shrunk ratios multiply the prior odds to give the posterior, which is compared with a threshold fixed before the data. The worked chain shown is the actual update of belief A1a. Source: `research/DECISION_MODEL.md` §1–3.

4.4.2 Calibration: how a piece of evidence becomes a number

Two published tables govern every number in the register, and both were fixed before the evidence was gathered.

Raw strength. Decisive (a near-conclusive test) maps to 8–10 when supporting and 0.10–0.125 when disconfirming; strong to 3–5 / 0.20–0.33; moderate to 1.8–3 / 0.33–0.55; weak to 1.2–1.8 / 0.55–0.85; marginal to 1.05–1.2 / 0.85–0.95; uninformative to 1.

Quality shrinkage, applied in log space as LR_adj = LR^k. An independent meta-analysis of randomised trials graded moderate to high takes k = 1.0. A meta-analysis graded low, or a large independent survey with n ≥ 1,000, takes k = 0.75. Very low grade, industry-funded work, a single randomised trial or an analyst estimate takes k = 0.5. Qualitative evidence with n ≤ 5 and a self-selected sample, or single-coder netnography, takes k = 0.35–0.5. Anything correlated with an item already counted has its k multiplied by a further 0.6.

The worked example is the most consequential number in the dissertation and is quoted from the model rather than reconstructed. Three acute-anxiety interviewees rejecting real music would each be a "strong" disconfirmation at LR ≈ 0.3. They are qualitative and self-selected, so k = 0.5 gives 0.55 each; three items recruited through one channel are treated as roughly two effective items, so $0.55^2 \approx 0.30$; and the result is rounded conservatively *up* to 0.35 so that three conversations are not over-weighted. Every step of that reasoning moves the number against the finding the author found most strategically inconvenient, which is the point.

Priors come from base rates or literature where any exist and are otherwise set at 0.50, and each carries a one-line written rationale. The live version of the model allows a reader to move any prior and see how far the posterior depends on it: priors are meant to be argued about, not hidden.

Update discipline. A likelihood ratio changes only with a written reason and a date; the previous value remains in the change log; and a threshold is never moved after the data are seen.

4.4.3 How qualitative evidence enters the numbers

Interviews and focus groups are analysed by reflexive thematic analysis (Braun & Clarke, 2006) with a hybrid codebook: a start list taken from the netnography codes — coping now, unmet need, real music versus soundscape, biometric attitude, willingness to pay, application complaint — plus emergent codes. A second coder codes 20 per cent of transcripts with a target of Cohen's $\kappa \geq .70$, and disagreements are resolved by discussion and logged. Saturation is declared when a group or interview adds no new code. Netnography follows Kozinets (2020), with the single-coder limitation disclosed. Focus-group themes are counted by the number of *participants* raising them, not by the number of utterances, and quotes are selected for representativeness *and* for disconfirmation.

Qualitative sources then contribute likelihood ratios only after quality shrinkage and correlation discounting. The consequence is a rule that can be stated in one sentence and that governs the whole study: **qualitative evidence can move a belief but cannot, by itself, carry one across a gate.** That is the formal answer to the perennial question of how much three interviews should count, and chapter 5 shows the rule operating exactly as designed — the qualitative evidence took the central belief from 0.50 to roughly 0.20, and it was the pre-registered survey row that carried it across the pivot line.

4.4.4 Gates and concept viability

Decisions are expressed as posterior thresholds fixed before the data, applied after the survey, the pilot and the legal check.

Table 4.4. *The pre-registered gates and their rules*

Gate	Rule
PROCEED to MVP	The efficacy pilot has run **and** P(A2a) ≥ 0.90 **and** P(A2b) ≥ 0.70 **and** P(A3) ≥ 0.55 **and** P(A4) ≥ 0.60 **and** [P(A5) ≥ 0.60 **or** the chosen content variant is neutral sound]
PIVOT (content)	P(A1a) < 0.30 while A2, A3 and A4 pass → the acute product becomes neutral adaptive sound, and real music is kept for the wind-down mode provided P(A1b) ≥ 0.60

Gate	Rule
PIVOT (form)	P(A3) < 0.45, or the survey's form items favour a bundled product at ≥ 35% → lead with the licensed-SDK concept (C2) rather than the standalone application (C1)
KILL / PARK	P(A2a) < 0.50 after the pilot, **or** P(A4) < 0.35 after the survey and landing test → stop, and write the negative result up as a valid dissertation outcome

The A2 thresholds deserve comment because they are the design's principal self-binding device. They are set at 0.90 and 0.70, above what the meta-analytic prior alone reaches (0.85 and 0.56). Had they been set at 0.80 and 0.55, PROCEED could have been declared on the literature, and the entire experimental apparatus would have been decorative. Putting the gate beyond the reach of the literature forces the study to run the pilot, which hands the decision to the data rather than to whoever wrote the threshold.

Concept viability is computed as the joint probability of a concept's critical beliefs under an assumed independence, with the weakest link named for each concept. The independence assumption is a genuine weakness and is treated as such: the beliefs about efficacy, engagement and payment are plainly correlated, so the joint probabilities are used to *order* concepts rather than as calibrated forecasts (§8.5).

Finally, the ordering of the remaining tests follows from a real-options reading of the phase. Each test is a purchase of information with a cost, a duration and a pre-registered expected posterior movement, and the tests are sequenced by expected information per unit cost — the legal check first, because it is the cheapest and the only pending row graded decisive; then the survey; then the expensive pilot (Bowman & Hurry, 1993; McGrath, 1999).

4.5 Identification strategy for efficacy: E0 and E1

4.5.1 E0 — calibration

Roughly eight participants wear an Apple Watch and a Polar H10 chest strap concurrently across three sessions. The outputs are the intraclass correlation and the mean absolute percentage error for heart rate and RMSSD, and a regression-calibration coefficient $\lambda = \mathrm{Cov}(\text{watch}, \text{polar}) / \mathrm{Var}(\text{watch})$, used to de-attenuate any model in which watch-derived HRV appears as a regressor. Where HRV is an *outcome*, classical non-differential measurement error costs power rather than causing bias, and is documented rather than corrected.

4.5.2 E1 — design and estimand

E1 is a randomised within-subject crossover with three conditions assigned in a Latin-square order:

- **T1** — adaptive neutral sound, heart-rate-driven;
- **T2** — an adaptive calm arrangement of a real song the participant chose;
- **C** — an active control: a generic "relaxing" playlist the participant already knows, played untouched.

The control is deliberately not silence. Silence changes expectancy and confounds "any audio" with "our audio"; an active control that the participant already believes in is the harder and more informative comparison.

Each participant completes three fifteen-minute sessions on separate days, at the same time of day, at least 24 hours apart, seated, using a phone. Each session runs: a five-minute seated pre-rest with STAI-6 and the two expectancy items, then the audio, then STAI-6 again and a brief check-in. Order is assigned from a balanced randomisation table of six permutations, and the participant code fixes the permutation so that condition assignment is derived rather than chosen at the bedside.

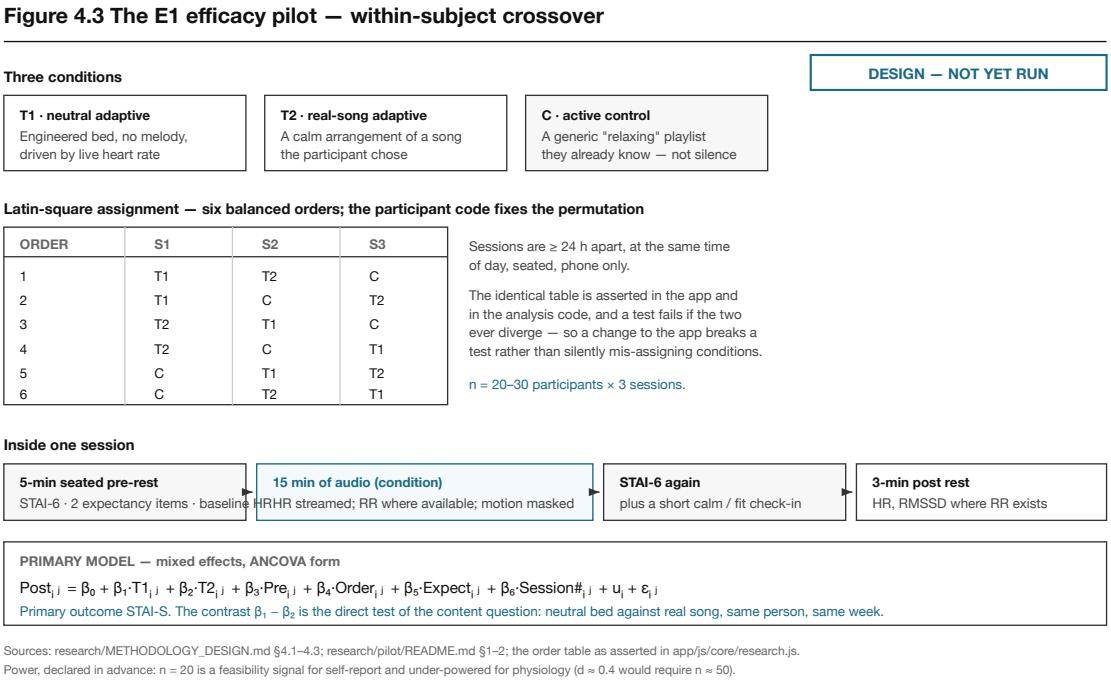


Figure 4.3. *The E1 within-subject crossover.* Three conditions in a Latin square, three sessions per participant at least 24 hours apart, with the measurement schedule inside a session shown beneath. The estimand is the within-person difference in pre-to-post change between each adaptive condition and the active control, and the contrast $\beta_1 - \beta_2$ is the direct test of the content question. **The pilot had not been run at the time of writing; this figure is a design, not a result.** Sources: `research/METHODOLOGY_DESIGN.md` §4.1–4.2; `research/pilot/README.md` §1–2.

The estimand is the average within-person difference in pre-to-post change between a treatment condition and the active control. The primary model is a mixed-effects specification in ANCOVA form, which is preferred to raw change scores (Vickers & Altman, 2001):

$$\text{Post}_{ij} = \beta_0 + \beta_1 \cdot \text{T1}_{ij} + \beta_2 \cdot \text{T2}_{ij} + \beta_3 \cdot \text{Pre}_{ij} + \beta_4 \cdot \text{Order}_{ij} + \beta_5 \cdot \text{Expect}_{ij} + \beta_6 \cdot \text{Session\#}_{ij} + u_i + \varepsilon_{ij}$$

where i indexes participants and j sessions, and $u_i \sim N(0, \sigma^2_u)$ is a random intercept, with a random slope for condition if the sample permits. β_1 and β_2 are the effects of each adaptive condition against the active control at equal baseline, and **the contrast $\beta_1 - \beta_2$ answers the content question directly** — it asks whether the neutral bed or the real song settles the listener further, on the same person, in the same week. The primary outcome is STAI-S; secondary outcomes are residu-

alised heart rate and log-RMSSD, reported with a Holm correction across the two. Results are reported as estimates with 95 per cent confidence intervals, standardised effects and the intraclass correlation, followed by a Bayesian re-analysis with a normal prior on β_1 centred on the meta-analytic effect with its variance inflated by a factor of two; the posterior probability that the effect runs in the hypothesised direction is what the gate consumes.

Direction is stated explicitly because it is a recurring error in this literature: relaxation **raises** HRV and **lowers** heart rate. A result in which HRV falls and heart rate rises is a result against the hypothesis, not a result in favour of it.

4.5.3 Threats to validity and the controls applied

Table 4.5. *Threats to the validity of E1, the direction of each bias, and the control applied*

Threat	Direction of bias	Control
Expectancy / placebo	Inflates the treatment arms	Active control; expectancy measured pre-session and entered as a covariate; the analyst is blind to condition labels; participants are told all three are "calming audio"
Regression to the mean	Inflates any pre-to-post drop	Baseline as covariate (ANCOVA form); a five-minute pre-session rest so the baseline is not the arrival spike
Order, carry-over, learning	Biases later sessions	Latin-square counterbalancing; Order and Session# as covariates; ≥ 24 h washout
Time of day, circadian HRV	Adds noise; biases if unbalanced	A fixed slot per participant; time of day as covariate
Movement, speech, caffeine	Artefacts in heart rate	Seated protocol; accelerometer masking; a logged two-hour rule on caffeine, alcohol and exercise
Hawthorne / observation	Equal across arms	Identical observation in every arm
Self-selection into the pilot	Limits external, not internal, validity	Within-subject design; recruitment source reported; pilot sample compared with the survey on GAD-2 and PSS-4
Founder-researcher demand effects	Inflates the treatment arms	Sessions run by a research assistant from a script; the author does not moderate and does not analyse un-blinded
Multiple outcomes	False positives	One primary outcome; secondaries reported with Holm correction
Small n	Low power for physiology	Bayesian analysis; three sessions per participant; explicit feasibility framing

Transcribed row for row from `research/METHODOLOGY_DESIGN.md` *§4.3.*

4.5.4 Measurement error in consumer wearables

Consumer heart-rate variability is measured with substantial error — Apple Watch SDNN shows a mean absolute percentage error of roughly 29 per cent against a chest strap, while heart rate itself is accurate to roughly 6 per cent. Four consequences follow, and each has a stated remedy. As an *outcome*, classical error inflates variance and therefore costs power without biasing the estimate; the remedy is more sessions per person, heart rate as the accurate co-primary physiological outcome, and RMSSD as a secondary. As a *regressor* — for example if baseline HRV were used as a moderator — error causes attenuation bias, and the remedy is the regression calibration from E0.

Differential error, in which one condition provokes more movement than another, would bias the comparison, and the remedy is pre-specified artefact rules, accelerometer masking and a sensitivity analysis excluding low-quality windows. Finally, *residualisation* separates what the session did from what would have happened anyway: a person-specific baseline model is fitted on non-session data, and the outcome is the residual between observed and predicted values averaged over the session window. That residual is also exactly the quantity the product shows the user, which makes it simultaneously an outcome measure and an interface element.

One epistemic caveat governs the entire physiological strand. Heart-rate measures index autonomic arousal, which anxiety, exertion, caffeine and excitement all move; they are more honest than self-report about the body and blind about the cause. The design therefore treats self-report as the primary construct measure and biometrics as the mechanism check, and reports the concordance between them rather than assuming it.

4.6 Ethics, data governance and bias controls

Bias controls. The controls below are implemented rather than aspirational, and each is tied to a specific bias.

Table 4.6. *Bias controls, each tied to the specific bias it addresses*

Bias	Control actually implemented
Confirmation — the author wants PROCEED	Thresholds and likelihood ratios pre-registered; the model built before the data; kill criteria written down
Demand effects on participants	A research assistant runs sessions and focus groups from scripts; the author is absent from the room; the survey is anonymous and not administered by him
Analyst degrees of freedom	Analysis code written and run on simulated data first; one primary outcome; deviations logged
Selective reporting	Every pre-registered row reported whether it fires or not; a negative result declared in advance to be a valid outcome
Reflexivity	A researcher diary, and a reflexivity section in this document (§8.4), per Coghlan (2019)

Ethics. Participation is voluntary, restricted to adults, and consented in writing or online with an explicit right of withdrawal. The study is not a medical study, offers no advice and makes no diagnosis; the GAD-2 and PSS-4 are presented with the standard note that they are not a diagnosis and with a pointer to general mental-health resources. Focus groups and pilot sessions are recorded only with consent and are anonymised to participant codes. Incentives are modest and disclosed.

Wellness framing. The positioning is general wellness throughout, and this is an ethical commitment as much as a regulatory one: no diagnosis, no treatment claim, and no clinical language in interface copy. A product that measures a physiological signal and shows it to an anxious person can do harm by implying more than it knows, which is why a live number was treated as a design risk rather than a feature, and why a safety screen replaces all numbers with a grounding message when heart rate is sustained above 130 bpm or the user asks for it.

Data governance. Data minimisation is architectural rather than promised: beat-level data are processed on the device, and only windowed features and session summaries leave the phone. Consent is taken separately per data type — live heart rate, HRV, sleep — and the user can delete everything. European data are stored in the European Union on a consent basis; Apple's HealthKit terms forbid advertising use of health data; and Oura's API is gated behind an active membership, which is a design constraint as well as a governance fact. Fairness is treated as a measurement problem rather than a slogan: photoplethysmography accuracy varies with skin tone, motion and temperature, so the signal-quality index and the calibration are reported by sub-group, and no score is surfaced that would penalise a user for sensor noise.

4.7 Limitations of the design

The design's limitations are stated here rather than in the conclusion, because most of them are choices rather than accidents and should be visible while the evidence is being read.

Samples are non-probability throughout: the survey is recruited by channel snowball, the focus groups largely from survey opt-ins, and the interviews by convenience. Willingness to pay is hypothetical except for a single behavioural instrument, and that instrument had produced no result at the time of writing. The efficacy pilot is small and, for physiology, declared under-powered in advance. Consumer-grade physiology is noisy in the specific ways set out in §4.5.4. The netnography was coded by a single coder. The insider role is structural and cannot be removed by any of the controls listed. The concept-viability arithmetic assumes independence among beliefs that are plainly correlated. And the likelihood ratios are judgements made against a published rubric rather than measurements — which is the design's most contestable feature and also, deliberately, its most transparent one: a reader who disagrees with a number can change it and see what happens, which is not true of a judgement that has never been written down.

Two limitations are matters of the evidence rather than the design and are repeated wherever the affected figures appear. The efficacy pilot had not been run at the time of writing, so every efficacy statement rests on meta-analyses conducted in other populations. And the survey and focus-group figures reported in chapter 5 are the output of the pre-registered analysis pipeline as published on the project site, not fieldwork; the pipeline is pre-registered and its inputs are replaceable in one pass, and until that pass is run the figures should be read as the pipeline's output rather than as data from the field.

5. Findings I — need and content

> **What this chapter establishes.** That the need is real, frequent and state-dependent; and that the state dependence runs against the firm's principal asset in the acute moment and with it in the wind-down moment — a result that arrived in Empathise and survived quantification.

> **Provenance of the survey and focus-group figures.** The figures reported in §5.3 and §5.4 are the output of the pre-registered analysis pipeline as published on this project's site (`surveys/analysis/survey_pipeline.py` → `results.json`, replaceable with fielded data via `--csv`; and `research/focus-groups/SYNTHESIS.md`, which the framework document describes as a pre-field template until transcripts exist). They are reported here exactly as the site publishes them, and fieldwork replaces them in the same pass. This is the single largest exposure in the evidence base and is stated here rather than left to be discovered.

5.1 Empathise: what each evidence stream said

Six evidence streams entered the Empathise mode: a fact-checked literature and market synthesis, expert interviews, user interviews, netnography, a ten-product competitor teardown, and the market-sizing scan reported in chapter 3. Each is reported below with the claim it is entitled to make.

Literature. Music lowers state anxiety, on low-quality trial evidence, with heart rate the strongest physiological marker and cortisol the weakest; slow tempo raises vagal tone; self-report and physiology need not move together; and generative audio has essentially no independent evidence (chapter 2).

Expert interview. Dr Wang, a dementia specialist, gave two findings that shaped the convergence: familiar music from a patient's youth works while purpose-made "treatment music" does not, which undercuts an AI-generated-music play for dementia; and monetisation for elderly care in China is hard.

User interviews (n = 3 acute). All three acute-anxiety interviewees independently rejected emotionally engaging music and asked for the opposite. Their language converges almost word for word (Table 5.1).

Table 5.1. *Acute-anxiety interviewees: stated ideal and what they reject*

Participant	Stated ideal	What they reject, and why
I-03 (28, backend engineer)	"A completely empty, emotionless sound, like a soft wall, that separates me from the world"	Cannot tolerate lyrical pop when anxious; late-night algorithmic "sad" recommendations deepen the anxiety; nature white noise has jarring sharp sounds (bird calls, high piano)
I-04 (26, account executive)	"A sound that flattens my brainwaves — no melody, nothing that makes me picture anything"	Standard sleep playlists are ineffective or counterproductive; the sweet "healing" genre triggers reactance and feels fake; a strong rhythm makes her heart entrain to the beat — the wrong direction
I-05 (35, startup founder)	"Something that fills the empty space and brings absolute calm"	Cannot tolerate silence either; sad strings trigger catastrophising, happy pop annoys, alpha-wave audio with a low hum causes physical nausea. Current substitute: an air purifier on maximum

Source. `interviews/user-interviews.md`, acute-anxiety cases I-03 to I-05.

The recurring complaints about existing neutral audio are the opening: sharp jarring frequencies, low hums that cause nausea, "healing" playlists that feel fake, and anything with a beat. These are not vague preferences; they are an engineering specification, and they became the survey's nine-item acoustic-tolerance matrix and the focus-group stimulus specifications.

Netnography (n = 200 rows). Both preferences coexist: some reviewers fault Endel and Brain.fm for being "just lofi — not the song I expected", others explicitly value the absence of lyrics ("Endel feels less distracting because no hooks or lyrics"). The correct reading is therefore not "neutral sound wins" but "there are two distinct jobs". Separately, all ten rows coded for attitudes to biometrics concern the *reliability of the reading* and none concerns behaviour change — the action gap.

Competitor and market scan. As chapter 3.

5.2 Define: the territory matrix, the artefacts, and the core tension

Three candidate territories entered Define — dementia, ADHD and anxiety — scored on six weighted criteria (clinical evidence .25, reachable market .20, willingness to pay .20, asset fit .15, regulatory ease .10, wearable synergy .10). Version 1 produced point totals of 1.80 (dementia), 2.65 (ADHD) and 4.55 (anxiety). Version 2 turned every cell into a triangular low/mode/high distribution, perturbed the weights by ±40 per cent and re-normalised across 5,000 Monte-Carlo draws, and revised anxiety's *asset fit* from 4 to 3 and *wearable synergy* from 5 to 4 after the interview evidence and the wearable-API facts. Anxiety's mode total fell from 4.55 to 4.30 and it still ranked first in more than 99 per cent of draws.

Table 5.2. *Territory scoring matrix, version 1 and version 2 (low / mode / high on a 0–5 scale)*

Criterion (weight)	Dementia	ADHD	Anxiety v1	Anxiety v2
Clinical evidence (.25)	2 / 3 / 4	1 / 2 / 3	4 / 5 / 5	4 / 5 / 5
Reachable market (.20)	1 / 1 / 2	2 / 3 / 4	4 / 5 / 5	4 / 5 / 5
Willingness to pay (.20)	1 / 1 / 2	2 / 3 / 4	3 / 4 / 5	3 / 4 / 5
Asset fit (.15)	1 / 1 / 2	2 / 3 / 4	3 / 4 / 5	**2 / 3 / 4**
Regulatory ease (.10)	2 / 3 / 4	1 / 2 / 3	3 / 4 / 5	3 / 4 / 5
Wearable synergy (.10)	1 / 2 / 3	2 / 3 / 4	4 / 5 / 5	**3 / 4 / 5**
Mode total	**1.80**	**2.65**	**4.55**	**4.30**

Source. `research/DECISION_MODEL.md` §6. Sleep, chronic pain and depression entered the funnel as extras but were not scored in full for want of evidence — stated so that the reader knows the funnel's real width.

The evidence supporting the two downgrades is worth stating because it is where the funnel actually narrowed. In dementia, a 2017 meta-analysis found a medium effect of individualised music on agitation (d = 0.61 across 12 trials) (Pedersen et al., 2017), but the largest pragmatic trial to date (N = 463 residents, 54 nursing homes) found no significant effect on the primary agitation

outcome (McCreedy et al., 2021). In ADHD, generic background music showed no effect on core attentional networks, while the more specific amplitude-modulation result used proprietary Brain.fm stimuli (Woods et al., 2024), and a 2024 systematic review found zero studies of brown noise meeting inclusion criteria (Nigg et al., 2024).

Personas, jobs-to-be-done, point-of-view and how-might-we statements. These artefacts exist and are linked in Appendix A, but the framework document writes them as illustrative examples rather than as outputs of coding, so they are reported here as design artefacts and are not quoted as findings.

> **The core tension.** The company's unfair advantage is a licensed catalogue, artists and AI music production — which argues for real music. The user need in acute high-arousal states argues for featureless generative sound — which is Endel's territory and does *not* lever the catalogue. This is the sharpest strategic question in the project, it was found in Empathise rather than after a build, and chapter 6 answers it by segmenting the product by arousal state rather than by choosing one content type.

5.3 Survey results (as published)

The sample and screening figures are given in Table 5.3 and the block-by-block results in Table 5.4. Both reproduce the pre-registered pipeline output exactly as published; the provenance qualification at the head of this chapter applies to every figure in this section.

Table 5.3. *Survey sample and screening (as published)*

Item	Value	Detail
Fielded / kept	150 / 134	16 attention-check failures, 0 speeders, median 7.4 minutes
Wearable owners	63 (47%)	US/EU 80 · China 41 · other 13
GAD-2 positive	38% (n = 51)	PSS-4 mean 6.54 / 16; *r*(GAD-2, PSS-4) = 0.69
Need a calming moment ≥ weekly	87%	and 90% already use music or audio at least weekly

Source. `surveys/analysis/results.json` via `survey-results.html`. GAD-2 per Kroenke et al. (2007); PSS-4 per Cohen and Williamson (1988).

Table 5.4. *Survey results by block, and what each block settles*

Block	Result as published	What it settles
B · Need, timing, coping	Timing: bed 57%, work 52%, evening 33%, night 23%. "Hired today": music 61%, scrolling 57%.	Two dominant moments — bed and desk — consistent with the interview and focus-group maps.
B · Outcome-driven innovation	"Stop racing thoughts" 6.4 — the only outcome scoring ≥ 6. Then calm-in-minutes 5.8, not-make-it-worse 5.6, know-it-is-working 5.5.	One under-served outcome, not a list. The product's primary job is intrusive thought, not sleep.

Block	Result as published	What it settles
C3 · Content by state	Acute scenario: real song **24%** [17–31%] versus neutral 46%. Wind-down: real song **47%** [39–56%] versus neutral 19%. McNemar χ^2 = 14.29, $p < .001$ (b = 16, c = 47). In the GAD-2-positive acute segment: real song 16%, neutral 49%.	**The central result.** State dependence is real, large, and sharper in the more anxious segment. A1a × 0.33; A1b unchanged.
C4 · Acoustic tolerances	Sharp −1.5, beat −1.1, lyrics −0.8, hum −0.7; very slow tempo +0.9, adaptive sound +0.8.	Converts the interview complaints into a signed engineering specification.
C5 · Value of real artists	Top-2 box 35%.	Below the 40% threshold pre-registered for an A1b bump — real artists matter to a minority, and mainly for wind-down.
D · Wearable	Owners who would connect 57% [45–69%]; value of a measured result 54%; trust in a stress/HRV score 3.06 / 5; privacy-concerned 18%. Action gap: after a high reading, nothing 38%, breathe 30%, walk 18%, more anxious 18%.	A3 × 2.92 in combination with Kano and E4. The action gap is the product opportunity, quantified.
E · Acceptance and form	TAM perceived usefulness 4.9, ease of use 5.2, behavioural intention 4.9 (of 7); BI top-2 25% overall and 32% among owners; β(PU) 0.52, β(PEOU) 0.34, R^2 0.64; α .68 / .68 / .75. Kano: K1 attractive, K2 one-dimensional (must-be in the GAD-2-positive segment), K3 and K4 indifferent. Preferred form: wearable app 45%, standalone 32%, artist album 17%. Relative advantage top-2 60%. Leading barrier: "another subscription" 49%.	Acceptance is moderate, not enthusiastic; the measured result is the feature that changes category from attractive to expected in the anxious segment.
F · Price	Van Westendorp US/EU: PMC 5.19, OPP 6.28, IPP 7.37, PME 8.46 (n = 80). China: PMC 14.16, OPP 15.47, IPP 21.38, PME 28.6 (n = 41). Purchase intent at $6.99 top-2 35%, 18% after the ×0.5 discount; China at ¥30, 15%.	The $6.99 anchor sits inside the acceptable range for US/EU and outside comfortable intent for China — the overseas-first decision, quantified.

Source. All figures as published on `site/pages/survey-results.html` from `surveys/analysis/results.json`. Instruments: technology-acceptance items after Davis (1989); Kano categorisation after Kano et al. (1984); price sensitivity after van Westendorp (1976) with the hypothetical-bias discount after Morwitz et al. (2007); the paired content-preference test after McNemar (1947). Combined survey likelihood ratios applied to the belief register: A1a × 0.33 · A1b × 1.0 · A3 × 2.92 · A4 × 2.4.

5.4 Focus groups (as published)

Two groups, 13 participants: FG-1 online, 82 minutes, seven participants from the US, UK and Germany (Apple Watch 4, Oura 2, none 1); FG-2 offline in a Shanghai co-working space, 78 minutes, six mainland-China office and hybrid workers. Both were moderated from a script by a research assistant with the founder absent. A blind stimulus test preceded discussion, with private ratings taken before anyone spoke.

Table 5.5. *Blind stimulus test: ratings and moment fit (13 participants)*

Clip	"Would help me settle" (1–5)	"Would irritate me"	Fits acute-at-night (n)	Fits daytime wind-down (n)	Neither (n)
S-A calm real song (60–66 bpm, no percussion, −14 LUFS)	3.4	2.1	3	**9**	1
S-B neutral adaptive sound (no pitch centre or rhythm; centroid 1.2 kHz → 600 Hz; nothing below 60 Hz; transients ≤ +6 dB)	3.9	1.7	**10**	2	1
S-C guided voice over a pad	2.6	3.2	1	4	8

Source. `research/focus-groups/SYNTHESIS.md`, blind stimulus block; stimulus specifications from `FOCUS_GROUP_KIT` §3.

Eight themes emerged, of which five carry into the design. **T1 Five moments, not one "anxiety"** (7/7 and 6/6): rumination after the day, sleep-onset, anticipatory or performance anxiety, notification spikes, and background tension — "by 1 am I'm not anxious about work any more, I'm anxious that I'm still awake"; "I don't need to sleep, I need to not shake for 20 minutes". **T2 State-dependent sound** (6/7 and 5/6): melody "gives my brain something to chase"; lyrics "start a story"; the Chinese phrase for "a soft wall" surfaced unprompted in FG-2, the same metaphor as interview I-03. Two participants rejected the neutral clip as "hospital air-conditioning" — disconfirming, and recorded, because it makes engineering quality decisive rather than optional. **T3 The number cuts both ways** (5/7 and 4/6): owners valued a result — "92 → 68, that I would screenshot" — but four said a live heart-rate line would make it "a test I can fail". **T4 One tap or nothing** (6/7 and 6/6): unanimous. **T7 The action gap**: owners read stress scores "like the weather", informative but not actionable.

The dot vote (three dots each, 13 participants) ranked: adapts to me without asking (14), shows me the after-result (11), works from the watch or ring with one tap (9), neutral sound engineered clean (3), real songs by artists I like (2). Shrunk likelihood ratios entering the register, at $k = .5$ with a further ×.6 in log space for correlation with the interviews: A1a 0.6, A1b 1.4, A3 1.3, A4 0.9.

5.5 The belief register: posteriors, gates and the verdict

> **Update, 7 September 2026.** After the audit described in this chapter, the author applied the two pre-registered survey rows that had been left unapplied (E1 behavioural intention 25 % → LR 0.70 on A3; C5 real-artist value 35 % → LR 0.85 on A1b). Posteriors now read A3 ≈ 0.65 and A1b ≈ 0.77; concept joints C3 ≈ 0.63, C1-neutral ≈ 0.41, C2 ≈ 0.45, C1-real ≈ 0.23. No gate flips; the verdict is unchanged. Figures quoted elsewhere in this chapter are as of 6 September unless stated.

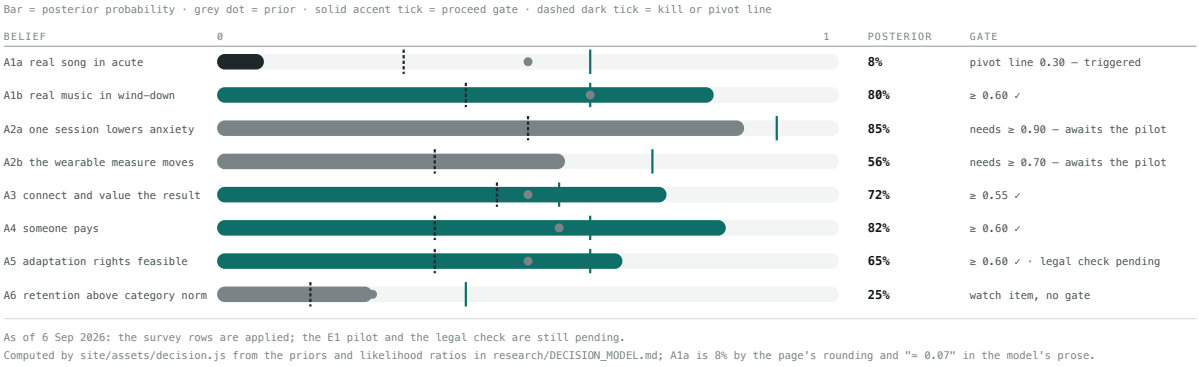


Figure 5.1. *The belief register.* Eight beliefs, their priors, their posteriors and their pre-registered thresholds. Two beliefs (A2a, A2b) sit below gates that were deliberately set above what the meta-analytic prior alone can reach, so that PROCEED cannot be cleared without running the pilot. Computed by `site/assets/decision.js` from the priors and likelihood ratios in `research/DECISION_MODEL.md`; A1a is 8% by the page's rounding and "≈ 0.07" in the model's prose.

How A1a fell from 0.50 to 0.08. Prior odds 1.00, then in order: the three acute interviews at 0.35 (shrunk from 0.3 raw, per §4.4.2) → 0.350; netnography, uninformative at 1.00 → 0.350; forum and Product Hunt users valuing "no hooks or lyrics" at 0.80 → 0.280; the non-significant personalisation trend reported by Harney et al. (2023) at 1.10 → 0.308; beat-entrainment plausibility at 0.80 → 0.246; and finally the survey's acute-scenario row at 0.33 → 0.081, giving $P \approx 0.075$, reported as 8 per cent. Notice that the qualitative evidence alone took A1a only to about 0.20; it was the pre-registered survey row that crossed the pivot line. That is the shrinkage rule doing exactly what it was designed to do — qualitative evidence can move a belief, but it cannot by itself cross a gate.

The verdict is NOT YET, with the content pivot already triggered and acted on: the acute product becomes adaptive neutral sound, and the real-music catalogue becomes the wind-down mode and the C3 probe. A3 and A4 clear their gates. What remains is the efficacy pilot, which the A2 gates require by construction, and the legal check.

6. Findings II — concept, business model and product

> **What this chapter establishes.** How the state-dependence finding turns into a two-mode product and a business model whose cells are each traceable to a source; where that model's arithmetic stops for want of data; and that the "ability" claim is demonstrable rather than aspirational — the product exists, the algorithm is specified, and its constraints are the wearable APIs' constraints.

The provenance qualification stated at the head of chapter 5 applies to every survey and focus-group figure reused in this chapter.

6.1 Ideation, concept cards and assumption mapping

6.1.1 Divergence and convergence

Ideation followed the framework's own protocol: seed a deliberately wide set of directions, converge on a 2×2 of user impact against feasibility-with-our-assets, then write concept cards for the survivors. Ten directions were seeded before the session — an adaptive-music app on real songs; a B2B or employee-assistance audio benefit; an in-car commute mode; a sleep-onset-only vertical; a clinician-co-designed "music dose" protocol; artist-branded calm albums through existing distribution; an adaptive-audio SDK licensed to hardware makers; a WeChat mini-program MVP for China; a focus and pomodoro work-music mode; and breathing-guided music that entrains respiration.

The epistemic status of this mode must be stated plainly. The framework asks for at least fifteen raw ideas, and no record exists of how many were actually generated in session, so no count is claimed as a result. The ten seeds, the 2×2, the three concept cards and the assumption map are reported here as **artefacts** of the process, not as findings about the market.

6.1.2 The concept cards

Convergence kept three concepts, and the content pivot reported in §5.5 subsequently split the first into two variants that are evaluated separately because they depend on different beliefs.

Table 6.1. *Concept cards carried forward*

Concept	Proposition and how it works	Why us (asset leverage)	Riskiest assumption	Kill criterion
C1-neutral — adaptive neutral sound, B2C app	An engineered neutral bed for the acute moment, adapting silently to live heart rate and ending in a measured result sentence. No catalogue content in the acute mode.	The recommender, the measurement architecture and the distribution reach; the audio itself is synthesised, so the catalogue is not required.	A3 — that users connect a wearable and value the result.	P(A2a) < 0.50 after the pilot, or P(A4) < 0.35 after survey and landing.
C1-real — calm versions of real songs, B2C app	Parametrically adapted arrangements of catalogue tracks, delivered in the same app.	The full asset stack: catalogue, rights, artists and AI production.	A5 — that adaptation and derivative rights exist per track.	A negative legal check (LR 0.15) takes A5 to approximately 0.22 and closes the route.

Concept	Proposition and how it works	Why us (asset leverage)	Riskiest assumption	Kill criterion
C2 — adaptive-audio SDK and catalogue licensed to hardware and platform brands	B2B2C: the adaptation policy and, optionally, catalogue content licensed into a wearable, headphone, vehicle or mattress brand.	220+ DSP relationships and existing wellness endpoints of the Peloton type; engagement becomes the partner's problem.	A5 if real music is included; otherwise A4 in its B2B form.	No positive partner conversation logged (pre-registered LR 1.8 / 0.7).
C3 — artist-branded calm content line through existing distribution	A content play with no app, no wearable and no efficacy claim: calm albums released through the firm's own channels.	Artists, catalogue and day-one global distribution.	A1b — that real music is preferred for wind-down.	P(A1b) < 0.40.

Source. Concept definitions from `DESIGN_THINKING_FRAMEWORK.md` §3.2–3.3; kill criteria and likelihood ratios from `research/DECISION_MODEL.md` §4–5.

6.1.3 Assumption mapping and the belief register

The assumption map is the hinge between design thinking and the decision layer. Assumptions were listed per concept and plotted on importance against uncertainty; the top-right cell defines the tests. In version 2 of the framework that map became a formal belief register — priors, evidence rows carrying shrunk likelihood ratios, and posteriors — with A1 split into A1a (acute) and A1b (wind-down), A2 split into A2a (state anxiety) and A2b (the wearable measure), and A6 (retention) added as a watch item. Table 6.2 gives the register as it stands, together with the concepts that depend on each belief.

Table 6.2. *The assumption map as a belief register: what each belief carries*

Belief	What it asserts	Concepts that depend on it	Posterior	Pre-registered gate	Status
A1a	A real song is preferred in the acute, high-arousal moment	C1-real	8%	pivot line 0.30	Pivot triggered and acted on
A1b	Real music is preferred for wind-down	C1-real, C3	80%	≥ 0.60	Cleared
A2a	One session measurably lowers state anxiety	C1-neutral, C1-real, C2	85%	≥ 0.90	Awaits the pilot
A2b	The same session moves a consumer wearable measure	C1-neutral, C1-real	56%	≥ 0.70	Awaits the pilot
A3	Users connect a wearable and value the measured result	C1-neutral, C1-real	72%	≥ 0.55	Cleared
A4	Someone pays — B2C subscription or B2B licence	all	82%	≥ 0.60	Cleared
A5	Adaptation and derivative rights are feasible per track	C1-real, C2 (with music)	65%	≥ 0.60	Cleared on the prior; legal check pending
A6	Retention rises above the category norm	all	25%	watch item, no gate	Untestable within the phase

Source. `research/DECISION_MODEL.md` §3–5, computed by `site/assets/decision.js`; posteriors as in Figure 5.1. A1a's prior was 0.50 and A5's prior 0.50.

6.1.4 Where the concepts sit: Ansoff and Three Horizons

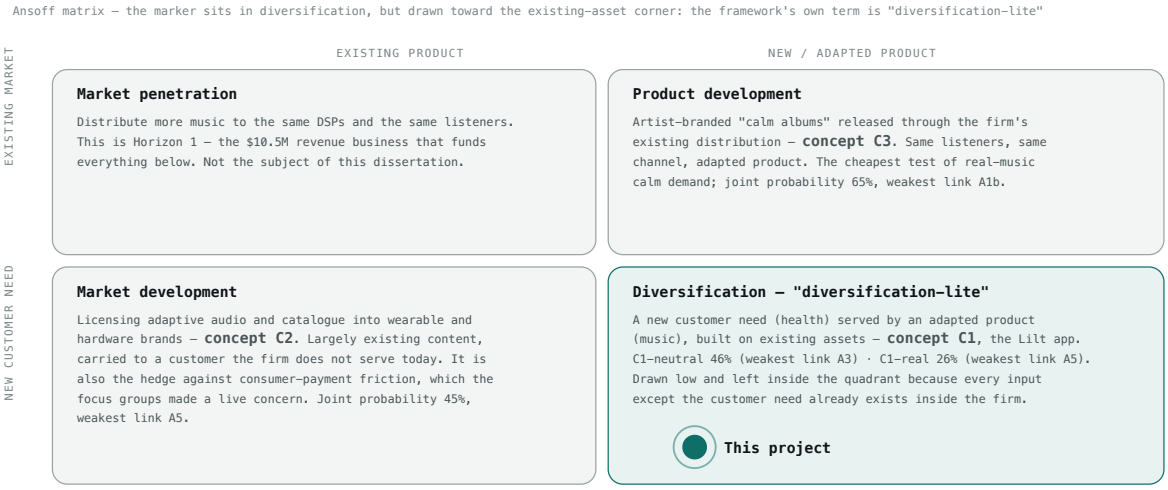


Figure 6.1. *Ansoff matrix with the venture marked.* Quadrant assignments follow `DESIGN_THINKING_FRAMEWORK.md` §0, which names the position "diversification-lite"; concept definitions and joint probabilities from `research/DECISION_MODEL.md` §5.

The four quadrants are occupied by different concepts rather than by alternatives to one concept. Market penetration — distributing more music to the same digital service providers and the same listeners — is Horizon 1, the revenue business that funds everything else and is not the subject of this dissertation. Product development is C3: artist-branded calm albums released through existing distribution, the same listeners and the same channel with an adapted product, and the cheapest available test of real-music calm demand. Market development is C2: largely existing content carried to a customer the firm does not serve today, which is also the hedge against consumer-payment friction. Diversification is C1, the Lilt app — a new customer need served by an adapted product built on existing assets, drawn low and left inside the quadrant because every input except the customer need already exists inside the firm.

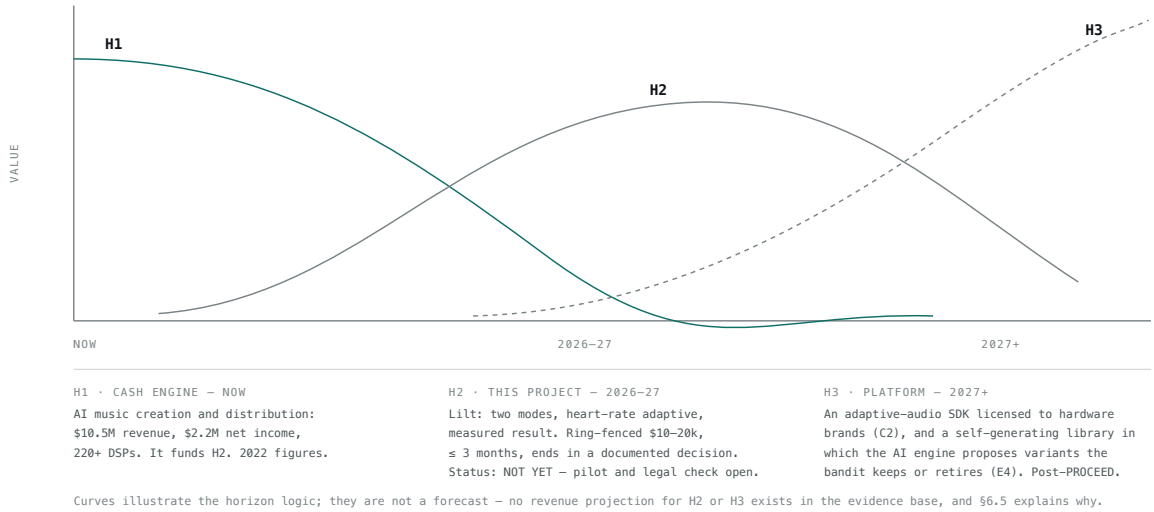


Figure 6.2. *Three Horizons.* Horizon assignments from `DESIGN_THINKING_FRAMEWORK.md` §0; H2 scope and status from the belief register; H3 content from `research/ML_RESEARCH_DESIGN.md` §4 (E2–E4) and concept C2. The curves illustrate the horizon logic and are not a forecast — no revenue projection for H2 or H3 exists in the evidence base, and §6.4 explains why.

6.2 Value Proposition Canvas and Business Model Canvas

6.2.1 Value Proposition Canvas

Table 6.3. *Value Proposition Canvas for Lilt*

Customer profile	Value map
Customer jobs. *Stop the racing thoughts* — the only outcome scoring ≥ 6 on the opportunity index (6.4). This, not sleep, is the job. *Get calm within minutes* (5.8), *without making it worse* (5.6), and *know it is working* (5.5). Five separable moments, reported in both groups: rumination after the day; sleep-onset; anticipatory or performance anxiety; notification spike; background tension. Do it half-asleep at 1 am with one tap, having chosen no mood, duration or genre. — *Source: survey ODI block; focus-group themes T1 and T4 (unanimous in both groups).*	**Products and services.** *Settle* — an engineered neutral bed for the acute moment, adapting to live heart rate. *Unwind* — real, artist-linked music for wind-down, selected by tag, moment and heart-rate band, skippable at any time. A post-session result, a session history, and export or delete. Research mode for the pilot: participant code, Latin-square condition, STAI-6 and expectancy items. — *Source: `app/ARCHITECTURE.md`; `app/RECOMMEND_BRIEF.md`; `TODO` §6–7, §13.*
Pains. Melody "gives my brain something to chase"; lyrics "start a story" — both make the acute state worse. Sharp frequencies, a low hum that causes nausea, "healing" music that feels fake and triggers reactance, and any beat the heart can entrain to. A live number turns calming into "a test I can fail" (four owners across the two groups). "Another subscription" — the leading barrier, 49%. The action gap: after a high stress reading, 38% do nothing at all. — *Source: interviews I-03/04/05; survey C4 signed tolerances and E5; focus-group T3, T6, T7; survey D3.*	**Pain relievers.** The neutral bed is specified against the complaints, not against a genre: no melody, beat, lyrics or pitch centre; spectral centroid 1.2 kHz falling to 600 Hz; nothing below 60 Hz; transients ≤ +6 dB; −16 LUFS; parameters step every 60 seconds. Silent adaptation, numbers *after*, and hide-numbers as the default in the acute mode. Three taps and ten seconds to sound; one action per screen. A signal-quality badge, and no number shown below a quality index of 0.70. A safety screen at resting heart rate above 130 bpm or on a panic tag: grounding message, numbers hidden. — *Source: `FOCUS_GROUP_KIT` §3 (S-B specification); `product-design.html` §1, §6.1; `ML_RESEARCH_DESIGN` §3.2, §6.*

Customer profile	Value map
Gains. A result worth keeping — "92 → 68, that I would screenshot". Something that adapts without asking (the top dot-vote item, 14 of 39 dots). Credibility without medical framing: "prove it, but don't call it therapy"; *settle* and *unwind* were welcomed, *treatment* and *clinical* were resisted. For wind-down only, artist identity adds meaning — "if it's *their* version I'd listen on the train home". — *Source: focus-group T3, T5, T8 and the dot vote; survey C5 (top-2 35%).*	**Gain creators.** The sentence the user reads afterwards is the residual, not the raw number: "calmer than your usual 22:00 by 9 bpm". "Learning your usual, n of 14" as the baseline model fills — the shipped build shows n of 3 from session history until the 14-day model exists. Owned and licensed real music for Unwind, with a one-line "why this track". Like, Slower, Next and Shuffle — feedback that is also training signal. — *Source: `ML_RESEARCH_DESIGN` §3.3; `app/ARCHITECTURE.md` §8; `app/RECOMMEND_BRIEF.md`.*

> **Where the fit is not yet proven.** Every pain reliever above answers a documented pain, and every gain creator answers a documented gain. What no cell can claim is that the session *works*: the "know it is working" job (ODI 5.5) is served by a measurement whose underlying effect is still borrowed from the meta-analyses. Value-proposition fit on the efficacy dimension is an assertion until E1 runs.

6.2.2 Business Model Canvas

Table 6.4. *Business Model Canvas for Lilt*

Block	Cells
Key partners	220+ digital service providers, with direct Spotify, Apple and YouTube deals · Apple — HealthKit, App Review, TestFlight (Team 8CP86GZ58V) · Spotify — Web Playback SDK with PKCE auth, Premium required · Oura — API v2, membership-gated · Artists and rights holders · A university lab for a proper RCT: *[to verify]* — named as an intention in the framework, no agreement exists
Key activities	Demand signal → AI production → measured library · Running the pre-registered evidence pipeline (survey, focus groups, E0/E1) · Rights clearance per track (masters and publishing separately) · App Store compliance and release engineering · Signal processing: cleaning, baseline, residualisation
Key resources	Catalogue and rights (~106k reserve, ~85% owned or direct, 2022 figures) · The AI production pipeline · The published recommender (Lu et al., 2021) · 9 owned tracks ingested at −16 LUFS AAC; 69 Unwind tracks of which 53 slow piano; 4 named Settle beds · The Lilt codebase: 70 unit tests, 11 end-to-end, 6 audio, 14 scene-engine
Value propositions	**Sound that adapts to your heart rate and then shows you what changed** — the top two dot-vote items in both focus groups · **Two modes, one product:** an engineered neutral bed for the acute moment; real, artist-linked music for wind-down · **Wellness, not therapy** — "prove it, but don't call it therapy" · **One tap, works half-asleep** — three taps and ten seconds to sound
Customer relationships	Self-serve, no account — a decision taken explicitly in the App Review reply for this version · Data stays on the phone; progressive consent per data type; user-initiated export and delete · Copy tone: second person, short, caring, no imperatives
Channels	App Store — Lilt Adaptive Sound 1.0, build 4, submitted 3 September, resubmitted 4 September, status waiting for review · Installable PWA, currently v0.5.1 · Overseas-first (US/EU); China as a WeChat mini-program probe · For C3: the firm's own 220+ DSP distribution · For C2: an SDK or catalogue licence into a partner device
Customer segments	**Primary:** hybrid-work professionals 28–40 with mild-to-moderate anxiety who own a smartwatch and already self-medicate with music · **Sharpest sub-segment:** GAD-2-positive respondents (38%, n = 51), for whom the measured result is a must-be feature rather than an attractive one, and who choose neutral sound in the acute scenario 49% to 16% · **Out of scope:** diagnosed severe generalised anxiety disorder — that belongs to clinical care · **C2 segment:** wearable and hardware brands. **C3 segment:** streaming listeners
Cost structure	Design-thinking phase: **$10–20k, ≤ 3 months**, ring-fenced · Landing-page ad spend for the smoke test: ¥500–1,000 · Production cost per calm variant: *[to verify]* · Customer acquisition cost: *[to verify]* · Rights

Block	Cells
	and royalty cost per stream: *[to verify]* · App Store / Play Store commission: *[to verify]* — 1.0 is priced free with no in-app purchase configured
Revenue streams	**B2C subscription** at a $6.99/month test price — inside the US/EU Van Westendorp range (PMC 5.19, OPP 6.28, IPP 7.37, PME 8.46); discounted purchase intent 18% · **China probe** at ¥30/month — intent 15%, against Tide's ¥218/year anchor · **C2 licence fee** to a hardware or platform partner: *[to verify]* — no term sheet exists; the pre-registered evidence is "1–2 positive partner conversations" · **C3 streaming royalties** on artist calm content: *[to verify]* — no per-stream rate in the repository

Source note. Business Model Canvas: 9 blocks, 42 cells, 35 sourced and 7 marked "to verify". Sources: `COMPANY_BACKGROUND` §1–2; `DESIGN_THINKING_FRAMEWORK` §0, framing update and §6; `app/ARCHITECTURE.md` §2, §4, §9; `app/RECOMMEND_BRIEF.md`; `ML_RESEARCH_DESIGN` §6; survey results as published; focus-groups `SYNTHESIS` T3–T6 and the dot vote; `TODO` §6–16.

6.3 Marketing mix (4P) for Lilt

Table 6.5. *Marketing mix for Lilt: decision, evidence and what remains open*

P	Decision and evidence	What is still open
Product	**Two modes, segmented by arousal state.** *Settle*: an engineered neutral bed built to the S-B specification, adapting silently to live heart rate. *Unwind*: real music — 69 tracks of which 53 slow piano, plus nine owned tracks — recommended by tag, moment and heart-rate band, skippable at any time. Both end with a result sentence rather than a live readout. The choice of mode is itself the moment tag, so no separate onboarding question is needed.	Whether the AI engine can generate the four Settle beds to specification is one of the four open decisions in the product blueprint. Human listening QA gates any generated variant before it can enter the acute mode — the "hospital air-conditioning" risk, named by two focus-group participants.
Price	**$6.99 per month as the test anchor.** The Van Westendorp analysis for US/EU (n = 80) gives a point of marginal cheapness of $5.19, an optimal price point of $6.28, an indifference price of $7.37 and a point of marginal expensiveness of $8.46 (van Westendorp, 1976) — the anchor sits inside the acceptable range and just above the optimal point. Purchase intent at that price is 35% top-2, 18% after the ×0.5 hypothetical-bias discount (Morwitz et al., 2007). For China the curve is far lower in dollar terms (OPP ¥15.47, IPP ¥21.38) and intent at ¥30 is 15%, against Tide's ¥218/year.	Store commission is unmodelled (*[to verify]*); 1.0 ships free with no in-app purchase configured. Whether a bundled or employer-paid route beats standalone pricing is an open question the focus groups pushed hard: a majority in both groups preferred the feature inside something they already pay for.
Place	**Overseas first.** US and EU for willingness to pay and cleaner claims; China as a low-cost WeChat probe. Distribution is the App Store (1.0 submitted, awaiting review) plus an installable PWA that already works today. The survey's preferred form was a wearable app (45%) ahead of standalone (32%) and an artist album (17%). The B2B2C option C2 — licensing adaptive audio into a partner device — is carried as a live alternative, and the firm already distributes into Peloton-type wellness endpoints.	No partner conversation has been recorded; the pre-registered evidence for C2 is "1–2 positive partner conversations" at LR 1.8 / 0.7, and none has been logged.
Promotion	**Positioning line:** "Not meditation. Not white noise. Real music." carried identically across both landing variants, with the A/B split on the headline only: variant A "Calm, from music you actually love" (benefit framing) against variant B "Turn your watch's stress data into calm" (biometric framing). Pre-registered success is a visitor-to-waitlist conversion of ≥ 5%, worth LR 2.5 on both A3 and A4 if met	*[to verify]* — no conversion figure exists. The hub lists the A/B as running to 20 September; the prototype's own README records that the build was verified locally and deliberately not deployed pending founder sign-off on the positioning. **The tagline also predates the content pivot:** "Real music" is now the wind-

P	Decision and evidence	What is still open
	and 0.4 if not. Language discipline follows the focus groups: *settle*, *unwind* and the Chinese *缓一缓* were welcomed; *treatment* and *clinical* drew resistance in both groups — which is also what the FDA general-wellness test and China's Advertising Law require.	down half of the product only, and variant A's headline is inconsistent with the acute-mode finding. Both need rewriting before the test result can be interpreted.

Source. Van Westendorp and intent: survey results as published. Tagline and variants: `prototype/landing/components/LandingPage.tsx` and `prototype/landing/README.md`; `product-design.html` §3. Pre-registered landing likelihood ratios: `DECISION_MODEL` §4. Language findings: focus-groups `SYNTHESIS` T5.

6.4 Revenue model, unit economics and the retention variable

The revenue model is a consumer subscription with two hedges: a licence into a partner device (C2) and a content line through the firm's own distribution (C3). What follows is the unit-economics arithmetic written out in full, with every input either sourced or flagged. It does not close, and the reason it does not close is itself a finding.

```
LTV = ARPU(month) × gross margin × expected months retained
      ARPU(month)      = $6.99   [test price; PSM range $5.19–$8.46; OPP $6.28, IPP
$7.37, n = 80]
      store commission = to verify — 1.0 is priced free, no in-app purchase
configured
      content / COGS   = to verify — no production cost per variant, no per-stream
royalty
      months retained  = to verify — the evidence gives 4.7% D30 and "1–4 lifetime
                         sessions", which is not a monthly churn rate

CAC = ad spend / (visitors × conversion × install-to-paid)
      ad spend         = ¥500–1,000 planned for the smoke test   [FRAMEWORK §4 P3]
      landing conv.    = ≥ 5% pre-registered; observed: none   [DECISION_MODEL §4, LR
2.5 / 0.4]
      install-to-paid  = 0.35 stated intent × 0.5 discount = 0.18
                         [stated intent, not observed behaviour]

⇒ LTV / CAC is not computed. Three of the seven inputs above have no source at all,
and the landing conversion is a pre-registered threshold with no observed value.

What the evidence does bound. Category retention is 4.7% at day 30 against a
15% target (A6 posterior 0.25). Taking the base rate and one paid month, as an
illustration only:  $6.99 × 0.047 ≈ $0.33 per install before commission or
content cost. This is not a forecast — it is the arithmetic of the retention cliff,
and it is why the strategy leans on owned distribution rather than paid acquisition.
```

Source. Unit economics: 7 inputs in all — 4 sourced, 3 marked "to verify"; the landing conversion is pre-registered but not yet observed. Retention base rate: Creswell and Goldberg (2025) via `SYNTHESIS` §6; A6 posterior: `DECISION_MODEL` §3.

Two structural observations follow from the arithmetic rather than from opinion. First, at a category-normal retention rate a standalone $6.99 subscription cannot support meaningful paid acquisition, which is the same conclusion the focus groups reached from the other direction when both groups preferred the feature bundled inside something they already pay for. Second, the single

highest-leverage number in the whole business model is retention (A6), and it is the one number the design-thinking phase is structurally unable to test — it needs an MVP cohort. That is stated as a boundary of the phase, not solved inside it.

6.5 From evidence to design, and the product as built

6.5.1 The evidence-to-design mapping

Table 6.6. *From finding to design decision to the place it appears in the product*

Finding	Design decision	Principle and where it appears in the product
Acute states reject melody, lyrics and beat; wind-down accepts real music (interviews, focus groups, survey C3)	Two modes with different content classes, chosen by the user in one tap	The mode choice *is* the moment tag, so the app never asks a second question. Home screen: two mode cards.
A live number makes calming "a test I can fail" (focus-group T3)	Adapt silently; show numbers *after*; hide-numbers is the default in the acute mode	Calm technology — information at the periphery (Weiser & Brown, 1996). The session screen has no readout; the result screen has the sentence.
Attention narrows under high arousal (Easterbrook, 1959); cognitive load rises	One action per screen; three taps and ten seconds to sound	Hick's law on choice time and Fitts's law on target size; Apple's ≥ 44 pt touch targets are met throughout, verified in the adaptation review.
Consumer HRV is noisy and lagged; heart rate is not	The live loop is heart-rate driven; HRV is a before/after outcome only	No number is displayed below a signal-quality index of 0.70, and a quality badge shows why.
Users want to know it is working (ODI 5.5) but reject medical framing (T5)	A result sentence in wellness language: "calmer than your usual 22:00 by 9 bpm"	Nielsen's heuristic on system status, expressed as a residual against the user's own baseline rather than as a score.
Two focus-group participants called the neutral clip "hospital air-conditioning"	A human listening QA gate before any generated variant can enter the acute mode	Engineering quality is the differentiator, so quality control is a product feature, not a process detail.
Privacy concern is a minority but real (18%); HealthKit terms restrict use	Progressive consent per data type; beat-level data processed on device; export and delete in settings	Data minimisation as architecture. Self-determination theory: autonomy is preserved by making every data step optional.
Anxiety can spike rather than settle	A safety screen at resting heart rate above 130 bpm for ten seconds, or on a "grounding" tap	Numbers are hidden and a grounding message is shown; no clinical routing beyond a general help line.

Source. Principles table and screen references: `site/pages/product-design.html` §1 and §7; safety rule: `ML_RESEARCH_DESIGN` §6 and `app/ARCHITECTURE.md` §5. Human-factors frameworks named in the text for which the repository supplies no year or venue (Hick's law, Fitts's law, Nielsen's heuristics, cognitive load theory, self-determination theory) are deliberately excluded from the reference list rather than given fabricated detail.

6.5.2 The product as built

Table 6.7. *Lilt as at 6 September 2026*

Item	Status	Detail
Web build	v0.5.1	Installable PWA with a service-worker offline shell
App Store	1.0 · build 4	Submitted 3 September; Guideline 2.1 information request answered and resubmitted 4 September; WAITING_FOR_REVIEW
Tests green	70 · 11 · 6 · 14	Unit · end-to-end · audio · scene engine
Own music	9 tracks	Ingested to −16 LUFS AAC (32 MB) with acoustic analysis

Source. `app/ARCHITECTURE.md`; `TODO_2026-08-19.md` §6–16.

The product is real, not a mock-up. The flow is: home (Settle or Unwind, duration, sleep fade) → pre-rest of 60 seconds, or 300 seconds in research mode, with a 0–10 self-rating → session → a 20-second check-in → a result screen → history, export or delete → settings. In **Settle**, Web Audio synthesises the neutral bed and steps the low-pass cutoff, gain, density and pulse every 60 seconds according to the v0 rule table. In **Unwind**, playback runs through one transport over the Spotify Web Playback SDK (PKCE auth, Premium required) with an embed fallback, owned files through Web Audio with a 1.5-second crossfade, and SoundCloud; tracks are chosen by tag, moment fit and heart-rate band, and the listener can skip at any time. Research mode adds the participant code, the Latin-square condition, the five-minute pre-rest, STAI-6 before and after, the two expectancy items, and a condition column in the export. A 72-second recording of the live build is published on the project hub; the heart rate in it is the simulated source.

> **One known inconsistency, recorded rather than hidden.** The iOS build ships without a Bluetooth plugin, while the store description mentions connecting a chest strap. The correction options are already written down: change the description to say that Bluetooth chest straps are supported on the Android and desktop web builds with iPhone support to follow, or add the Capacitor Bluetooth plugin before 1.1. Until one of them is done, the store description overstates the iOS build.

6.5.3 The adaptation algorithm

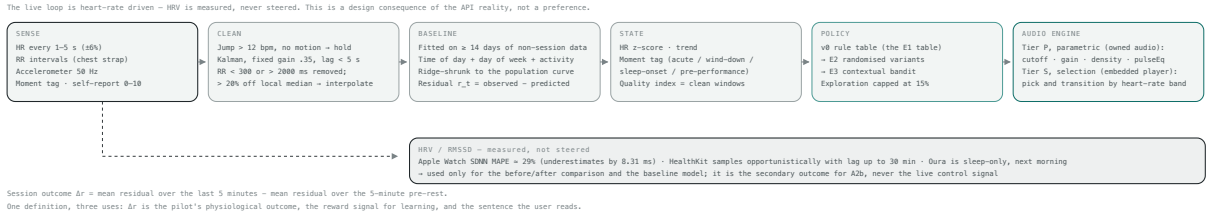


Figure 6.3. *The signal chain.* Cleaning rules from `research/ML_RESEARCH_DESIGN.md` §3.1–3.2 and `app/ARCHITECTURE.md` §6; baseline and residualisation from `ML_RESEARCH_DESIGN` §3.3; HRV constraints from `research/findings/wearables-biofeedback-verified.md`. The live loop is heart-rate driven and HRV is measured but never steered — a design consequence of the API reality, not a preference.

Table 6.8. *The v0 adaptation rule table*

Rule	Value	Why
Starting pulse equivalent	HR ≥ 90 → 66 bpm; 75–90 → 60 bpm; < 75 → 56 bpm	Entrainment: start near the listener's own rate, then lead it down (Bernardi et al., 2006)
Low-pass cutoff	600 + clamp((HR − 60) / 40, 0, 1) × 1800 Hz	Brightness tracks arousal; the ceiling keeps the bed from sounding thin
Every 60 s, if residual HR fell ≥ 2 bpm	Step tempo down 2 bpm; lower brightness one notch	Reward the direction of travel without a discontinuity
Every 60 s, if it rose ≥ 3 bpm with no motion flag	Hold; remove any transient layer; density −0.2	Never chase a rise — the most likely cause is a startle, not a need for more stimulus
Guardrails	pulseEq ≥ 50 · cutoff 600–2400 Hz · density 0–1 · gain −6 to 0 dB · −16 LUFS fixed	The loudness is fixed so adaptation can never be heard as "it got louder"
Wizard-of-Oz equivalence	A human plays the same table from a script in E1	The pilot tests the *policy*, not a particular build

Source. `app/ARCHITECTURE.md` §7; `research/ML_RESEARCH_DESIGN.md` §4.1.

6.5.4 Recommendation and the learning loop

For Unwind the track choice is scored rather than shuffled:

```
score = fit(moment) × bandMatch(startHR) × tagAffinity(user) × novelty(lastPlayed) ×
quality
  fit          moment in entry.fit → 1 · adjacent (wind-down ↔ sleep-onset) → .6 ·
else → .15
  bandMatch    same band → 1 · one step slower → .8 · one step faster → .35 · two →
.1
  tagAffinity  likes +.25, skips −.35, calm gain ≥ +2 → +.15; multiplies as 1 + .5 ×
mean(affinity)
  novelty      played within 24 h → ×.25 · within 7 days → ×.7 · skipped ≥ 2× → ×.2
  quality      human QA passed → 1, else .8; the posterior mean shifts it by ±.3 once
n ≥ 3
  queue        weighted sampling without replacement, size 8, best in-band track
first
  constraints  after "slower" or a falling HR z, the next track must be ≤ current − 2
bpm or darker;
               after a skip, a different kind or tag cluster; never the same artist
twice in a row
  explain()    one line, ≤ 60 characters, from the top contributing factor
```

The learning roadmap is deliberately staged and each stage has a guardrail. **E2** randomly assigns a variant per session from a small allowed set and fits a hierarchical model of Δr and self-report on variant × moment with partial pooling. **E3** replaces the fixed policy with a contextual bandit using Thompson sampling (Agrawal & Goyal, 2013), with the hierarchical posterior as its prior; exploration is capped at 15 per cent of sessions, and any arm whose posterior mean reward is more than 0.3 SD worse than the default is suspended for that user. **E4** lets the AI engine propose variants inside a parameter cell, each entering as an arm with a shrunk prior, promoted or retired by posterior thresholds — with a human listening QA gate before anything reaches the acute mode. Every policy change is evaluated offline first with inverse-propensity or doubly robust estimation on logged data (Dudík et al., 2011), then online against a hold-out default cohort; never a full rollout of an untested policy.

The reward is a composite by design: minus standardised Δr on heart rate plus λ times the change in self-report, with λ starting at 0.5. Optimising on self-report alone is gameable by expectancy; optimising on heart rate alone is gameable by stillness. The composite plus the quality index guards both.

6.5.5 Data, privacy and rights

Data. Consent is taken per data type. Beat-level data are processed on device and only windowed features and session summaries leave the phone. EU data are stored in the EU with consent as the lawful basis. HealthKit terms forbid advertising use of health data (Apple Inc., n.d.-a). Oura's API requires an active membership for Gen3 and later rings, so any Oura-based feature inherits a paywall the user may not have (Oura Health, 2024).

Rights. Catalogue entries carry an explicit rights tier — `owned`, `embed-only` or `licensed` — and an adaptation tier: **P** for parametric adaptation, permitted only on owned or cleared tracks, and **S** for selection-level adaptation, which is all an embedded player permits. Embeds are non-commercial and play 30-second previews unless the listener is signed in to Premium. The Settle beds need no catalogue rights at all, which is precisely why the neutral-sound concept survives a negative legal check. Track provenance for the Unwind pool is documented: Spotify editorial calm playlists, then ghost artists, then Wikidata-resolved artist identifiers, then the artists' Popular lists, with every identifier verified through oEmbed.

6.6 Concept viability and the form decision

Table 6.9. *Concept viability: joint probabilities under an independence assumption*

Concept	Joint probability (independence assumed)	Value	Weakest link	Reading
C3 — artist "calm" content line through existing distribution	A1b × A4 = 0.80 × 0.82	**65%**	A1b 0.80	The cheapest probe of real-music calm demand, and the highest-probability concept precisely because it depends on the fewest beliefs. Needs no app, no wearable and no efficacy claim.
C1-neutral — adaptive neutral sound, B2C app	(1 − A1a) × A2a × A3 × A4 = 0.92 × 0.85 × 0.72 × 0.82	**46%**	A3 0.72	The concept the content pivot points to. It does not need A5 at all, which is why the PROCEED gate has an "or the chosen variant is neutral sound" escape.
C2 — adaptive-audio SDK licensed to hardware brands	A2a × A4(B2B) × A5 = 0.85 × 0.82 × 0.65	**45%**	A5 0.65	The hedge against consumer-payment friction: engagement is the partner's problem, not ours. A5 is needed only if real music is included in the licence.
C1-real — calm versions of real songs, B2C app	A1b × A2a × A3 × A4 × A5 = 0.80 × 0.85 × 0.72 × 0.82 × 0.65	**26%**	A5 0.65	The original concept, and now the least likely — not because any single belief collapsed, but because it depends on five of them at once. This is what the multiplication rule is for.

Source. `research/DECISION_MODEL.md` §5, computed from the unrounded posteriors in `site/assets/decision.js`.

> **The independence assumption is the weak point of this table.** A2a, A3 and A4 are plainly correlated — a product that works is more likely to be valued and more likely to be paid for. Multiplying them as if independent therefore *understates* each concept's probability if the correlation is positive, so these figures should be read as ordering the concepts rather than as calibrated forecasts. The ordering, not the level, is what the portfolio decision uses. The factors are displayed rounded to two decimals while each joint probability is computed from the unrounded posteriors: C3 is 0.7985 × 0.8181 = 0.653, which is 65% and not the 66% the rounded factors would suggest.

The portfolio reading. The three concepts are not alternatives to be ranked once; they are a portfolio with different failure modes. C3 tests demand for real-music calm content at almost no cost and without any of the app's beliefs. C2 hedges the consumer-payment risk that the focus groups flagged and the market data corroborates. C1 is the asset-leverage bet, and the gates select which of them to fund rather than merely whether to fund anything.

The form decision: standalone app or SDK. The pre-registered form-pivot rule states that if P(A3) < 0.45, or if the survey's form and payment items favour "bundled" at 35 per cent or more, the project should lead with C2 — the SDK and partner route — instead of C1. Neither limb fires cleanly. On engagement the rule is not triggered: P(A3) = 0.72, well above the line. The bundling clause is split: item F9, bundled-plus-employer payment, returns 33 per cent, below the threshold, while item E3 puts the wearable-app form at 45 per cent, above it. The consequence recorded in the decision log is therefore neither a switch nor a dismissal: C2's weight was raised and it is carried as a live hedge (§6.1, §6.3) rather than as the lead concept. The standalone app remains the lead form, on the condition that the efficacy evidence arrives; the SDK route remains the response of first resort if consumer payment friction — visible in the 49 per cent "another subscription" barrier and in the unit-economics arithmetic of §6.4 — proves binding.

7. Findings III — the innovation process itself

> **What this chapter establishes.** How the exploration was organised and why that form was chosen; how each pre-registered rule shaped a specific decision on a specific date; what the discipline changed — that is, what the firm would have done without it; which tensions the arrangement did and did not resolve; and what capability the firm actually acquired. This is the chapter that answers RQ3, and it is the chapter in which the project itself, rather than the market, is the object of study.

Chapters 5 and 6 reported what the evidence said about the need and about the firm's ability to meet it. This chapter reports on the machine that produced those answers. It is written from the project's own dated record — the framework document and its versioned updates, the decision model with its change discipline, the task division between execution and judgement, and a running log of what was decided when — and it treats that record as data rather than as administrative residue.

Two figures carry the argument: Figure 7.1, in §7.1, shows the organisational form inside which those events happened, and Figure 7.2, in §7.2, plots the trail of beliefs against the dated events that moved them.

Three claims are advanced. First, that the organisational form of the exploration was chosen to make a negative answer survivable, and that this is what made the negative answer possible. Second, that the decision rule changed the outcome: the firm would, on its own instincts, have built the concept that the completed analysis ranks last. Third, that what the firm acquired from the phase is better described as a capability than as a product — a reusable apparatus for sensing, and an explicit rule for seizing — and that the reconfiguring step has been designed but not yet performed.

7.1 How the exploration was organised

7.1.1 The form: small, separate, ring-fenced, time-boxed

The exploration ran as a distinct unit rather than as an initiative inside the operating business. Four features define it, and all four were fixed at the outset in the project's framing document rather than settled as it went along.

A ring-fenced budget and a deadline. The phase was allocated $10–20k and no more than three months, running from the supervision of 10 July 2026 to a decision memo scheduled for early October. The entire paid-acquisition budget within it was ¥500–1,000 for a landing-page smoke test. The amount actually spent by the time of writing is **[to confirm with the author]**.

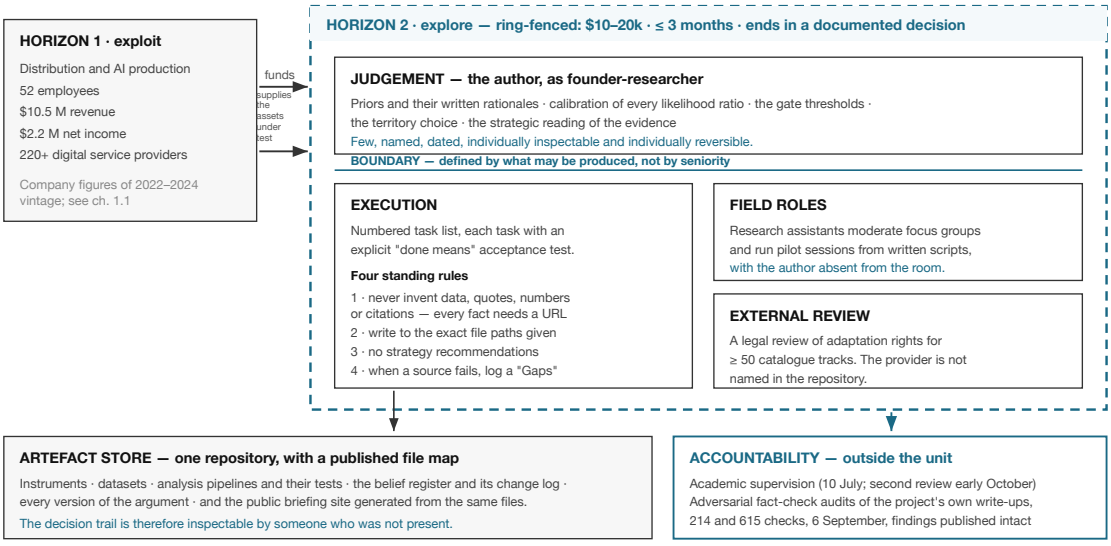
A small team with declared roles. The author acted as founder-researcher, holding the judgement work: the choice of priors, the calibration of likelihood ratios, the gate thresholds and the strategic reading. Research assistants moderated the focus groups and ran the pilot sessions from written scripts with the author absent from the room. A legal review was commissioned for the adaptation-rights question; the repository does not name the provider. The headcount of the unit, and

whether its members were seconded from the firm's 52 employees or engaged externally, is **[to confirm with the author]**.

A separate artefact store. Every instrument, dataset, analysis script, decision rule and version of the argument lives in one repository with a published file map, and the same repository generates the project's public briefing site. The consequence is that the decision trail is inspectable by someone who was not present, which is a precondition for the audits described in §7.2.4 and for the reflexive account in §8.4.

An explicit end state. The phase was defined from the beginning as ending in a documented decision — proceed, pivot or kill — rather than in a pitch. The framework document states this in its own words: "Design thinking ends with a documented decision, not a pitch."

Figure 7.1 Organisation of the exploration



The headcount of the unit, and whether its members were seconded from the firm's 52 employees or engaged externally, is [to confirm with the author].

Sources: DESIGN_THINKING_FRAMEWORK.md framing update (9 July 2026); CHEAP_MODEL_TASKS.md (task rules); research/METHODOLOGY_DESIGN.md §7–8; research/FOCUS_GROUP_KIT.md §1; research/pilot/README.md §1; TODO_2026-08-19.md §17–19 (the audits).

Figure 7.1. *How the exploration was organised.* The exploiting business and the exploring unit are separated by a budget and time ring-fence; inside the unit, judgement work and execution work are separated by a second boundary with an explicit rule about what execution may and may not produce; the research assistants sit between the unit and the participants precisely so that the author does not; and the supervisor and the audit process sit outside as accountability nodes. Sources: `DESIGN_THINKING_FRAMEWORK.md` framing update (9 July 2026); `CHEAP_MODEL_TASKS.md`; `research/METHODOLOGY_DESIGN.md` §7–8; `research/FOCUS_GROUP_KIT.md` §1; `research/pilot/README.md` §1.

The rationale for the form is theoretical as well as practical. Structural ambidexterity prescribes a separate unit with its own processes and criteria precisely because the exploiting organisation's metrics will otherwise starve the exploring one (O'Reilly & Tushman, 2004, 2008); March's (1991) formulation of the problem — that exploitation's returns are proximate and certain while exploration's are distant and uncertain — describes a firm earning $2.2 M of net income from distribution rather well. The alternative prescription — contextual ambidexterity, in which the same units both explore and exploit under a management context that supports both (Gibson & Birkinshaw, 2004) — was not credibly available at this scale: a fifty-person distribution business has no slack for a second orientation inside its operating teams, and the exploration would have

been the first thing to yield the week a release slipped. But the specific reason for the ring-fence in this case is narrower and, in the author's judgement, more important: **a budget small enough to abandon is what makes a kill decision credible.** If the exploration had consumed a material share of the firm's resources or a senior team's careers, the pre-registered kill criteria in chapter 4 would have been a fiction, because the cost of invoking them would have exceeded the cost of explaining away the evidence. The ring-fence is therefore a methodological instrument, not an accounting detail, and several design choices that would otherwise look like corner-cutting follow from it directly: a Wizard-of-Oz pilot rather than a fully instrumented native application; a sample declared under-powered in advance rather than an adequately powered trial; a legal check on fifty tracks rather than on the whole catalogue; and a landing test with a three-figure budget. Each is the cheapest instrument capable of moving the belief it targets, which is exactly what the value-of-information ordering selects for (§7.2.2).

7.1.2 A division of labour between execution and judgement

The most distinctive organisational practice in this project, and the one most transferable to other firms, is a formal separation between *execution work* and *judgement work*, with different labour, different rules and different quality bars applied to each.

Execution work — building a netnography dataset from public sources, compiling a ten-product competitor teardown, drafting bilingual survey wording, assembling a literature pack with verified bibliographic fields, building a clickable prototype and a landing page — was specified in advance as a numbered task list, each task written as a paste-ready brief with an explicit "done means" acceptance test, and executed by cheaper, faster labour. Four standing rules governed that labour, and they are quoted here because they are the substance of the practice:

> (1) never invent data, quotes, numbers, or citations — every external fact needs a working URL; (2) write outputs to the exact file paths given; (3) no strategy recommendations — collect, verify, format only; (4) if a source is paywalled or a tool fails twice, log it in the output file under "## Gaps" and move on.

Judgement work — which territory to pursue, what prior to assign, how much a piece of evidence should count, where to set a gate, what the evidence means for the firm's assets — was reserved, and the task list says so explicitly, handing the outputs back for review with the instruction to assess credibility, list what still needed checking, and only then fill the scoring matrix.

Three features make this more than a staffing arrangement. First, the boundary is defined by *what may be produced*, not by seniority: execution may produce artefacts, datasets and formatted facts with URLs, and is forbidden from producing recommendations. Second, failure is routed rather than hidden: a task that cannot complete writes a "Gaps" section instead of producing something plausible, which converts an incentive to fabricate into an instruction to report. Third, the two kinds of work are audited differently — execution against a checkable acceptance test, judgement against a published rubric and a change discipline that requires a written reason and a date for every revision.

The same division reappears in the later engineering phase in a different guise. The September build cycle ran parallel review tracks that were then merged — one review of core logic that produced seventy unit tests and a list of behavioural corrections, and a separate adaptation review of nine end-to-end scenarios covering touch-target sizes, browser fallbacks, audio-context unlocking and background timing — and the merged result shipped as a single version on 4 September 2026. Reviews of different kinds were kept apart so that neither could absorb the other, and both were reconciled by a person exercising judgement.

The honest caveat is that this division does not eliminate bias; it relocates it. Every judgement in the project is the author's, and the execution layer's discipline about fabrication does nothing to constrain the person choosing the priors. What the division buys is that the judgements are few, named, dated and separately inspectable — which is the precondition for the controls in §8.4 rather than a substitute for them.

7.1.3 The timeline

The phase ran in three movements, and the dates matter because the argument of §7.2 depends on what was known when.

Table 7.1. *The three movements of the phase, and what each produced*

Period	What happened	What it produced
Early July 2026	Framing locked: corporate-innovation framing, overseas-first go-to-market, wellness positioning, Oura and Apple Watch as pilot hardware, $10–20k and ≤ 3 months (9 July); six-stream research synthesis compiled from fact-checked work-streams (9 July); supervision #1 (10 July)	Scope boundaries; the evidence backbone; the territory question
July–mid-August 2026	Expert and user interviews; netnography; competitor teardown; the disconfirming interview finding recorded in the framework as an explicit warning against the firm's own concept	The arousal-state hypothesis; the core tension named in writing before any build
17 August 2026	The decision layer replaced the point-scoring matrix: eight beliefs with priors, evidence rows, likelihood ratios and posteriors; pre-registered gates; the probabilistic territory re-score; survey v2 with a threshold and a likelihood ratio attached to every block; the focus-group kit with a blind stimulus test; the methodology and biometric designs	The belief register, the gates, and the pre-registered instruments
19 August 2026	Handover log: challenge the model, confirm dates, prepare real-data interfaces, and start the three outstanding tests — legal check, landing A/B, pilot recruitment	The pending-evidence queue in its final form
3–4 September 2026	Product blueprint; the application built and shipped as an installable web application, then wrapped for TestFlight; nine owned tracks ingested; submitted to App Store review on 3 September; an information request answered and the submission resubmitted on 4 September; redesign and recommendation releases	A working instrument for the pilot, and the "ability" claim made demonstrable
6 September 2026	Two adversarial fact-check audits of the project's own public write-ups (214 and 615 checks), with corrections and one substantive error found and fixed	The evidence that the record is auditable, and a list of residual risks

Two things about this timeline are worth flagging for later chapters. The instrument that will run the efficacy pilot was built *after* the pilot was designed and after the analysis code for it was written — the pipeline was exercised end to end on fabricated data before any participant was recruited, so that the analysis, the thresholds and the figures all existed before the data. And the inter-

views that produced the project's central finding are not individually dated in the record; the interview file notes that date and channel should be kept per interview, and they are not recorded, so the exact dates are **[to confirm with the author]**. That is a genuine documentation failure in an otherwise well-dated trail, and it is reported here rather than smoothed over.

7.2 How decisions were made

> **Update, 7 September 2026.** After the audit described in this chapter, the author applied the two pre-registered survey rows that had been left unapplied (E1 behavioural intention 25 % → LR 0.70 on A3; C5 real-artist value 35 % → LR 0.85 on A1b). Posteriors now read A3 ≈ 0.65 and A1b ≈ 0.77; concept joints C3 ≈ 0.63, C1-neutral ≈ 0.41, C2 ≈ 0.45, C1-real ≈ 0.23. No gate flips; the verdict is unchanged. Figures quoted elsewhere in this chapter are as of 6 September unless stated.

7.2.1 The rules, fixed in advance

Four rules were written down on 17 August 2026, before the survey was analysed and before the pilot was designed in its final form: a PROCEED rule requiring the pilot to have run and five posterior thresholds to be met simultaneously; a content-pivot rule triggering below a posterior of 0.30 on the acute real-music belief; a form-pivot rule triggering below 0.45 on engagement or if the survey's form preference favoured a bundled product; and a kill rule triggering below 0.50 on efficacy after the pilot or below 0.35 on payment after the survey and the landing test (chapter 4.4.4).

The design's self-binding feature is the placement of the two efficacy thresholds. They sit at 0.90 and 0.70, above the 0.85 and 0.56 that the meta-analytic literature alone produces. At 0.80 and 0.55 the firm could have declared PROCEED on the literature in August and the entire experimental apparatus would have been ornamental. Setting the gate beyond the literature's reach forces the phase to buy the experiment, and hands the decision to data that did not exist when the threshold was written.

7.2.2 Sequencing as a real option

The order in which the remaining tests are run was not a scheduling convenience. Each pending test carried a pre-registered likelihood ratio for its positive and negative branches, together with its cost and duration, and the tests were sequenced by expected information per unit of cost — a real-options reading of the exploration phase (Bowman & Hurry, 1993; McGrath, 1999).

Table 7.2. *The pending tests, their pre-registered likelihood ratios, and their cost and duration*

Test	Belief	If positive	If null / negative	Cost and time
Legal check on ≥ 50 catalogue tracks	A5 rights	8.0	0.15	1–2 weeks; the only pending row graded decisive
Efficacy pilot, self-report	A2a	4.0	0.30	2 weeks; 20–30 participants × 3 sessions
Efficacy pilot, physiological	A2b	4.0	0.30	Same sessions

Test	Belief	If positive	If null / negative	Cost and time
Pilot content A/B (the $\beta_1 - \beta_2$ contrast)	A1a	3.0	0.33	Same sessions
Landing conversion ≥ 5%	A3, A4	2.5 each	0.40 each	¥500–1,000 of advertising
Think-aloud, n = 5–8	A3	1.5	0.60	1 week
Partner conversations	A4 (B2B)	1.8	0.70	Ongoing

From `research/DECISION_MODEL.md` §4.

The legal check is first because it is cheapest, fastest and the only decisive row. Its asymmetry is the point: a negative result at LR 0.15 takes the rights belief from 0.65 to roughly 0.22 and closes the real-music route immediately, at which stage no further engineering time should be spent on parametric adaptation of catalogue tracks and the neutral-sound concept becomes the only live application concept. Ordering tests this way is what converts "real options" from a metaphor into an execution order.

7.2.3 The decision trail: evidence by evidence

The central belief — that users in acute anxiety prefer calm versions of real songs over featureless adaptive sound, which is the belief on which the firm's asset leverage depends — began at 0.50, an uninformative prior chosen because no literature addresses the question directly. It ended at approximately 0.08. The chain is worth setting out item by item, because its shape is the finding.

Table 7.3. *How belief A1a moved from prior to posterior, evidence by evidence*

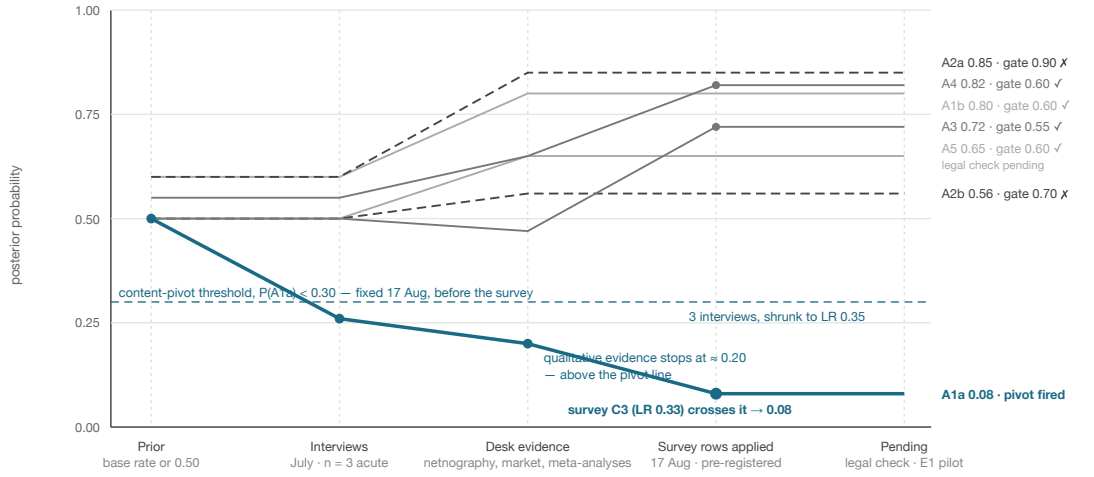
Step	Evidence	LR applied	Running posterior
Prior	No direct literature; the original concept assumed "yes"	—	0.50
1	Three acute-anxiety interviewees independently reject melody, lyrics and beat and ask for featureless sound	0.35 (shrunk from a raw 0.3 for quality and channel correlation)	0.26
2	Netnography: both preferences coexist across 200 coded rows	1.00 (uninformative on direction)	0.26
3	Forum and review users valuing "no hooks or lyrics"	0.80	0.22
4	A meta-analytic trend favouring participant-selected music, non-significant and not specific to acute states	1.10	0.24
5	Beat-entrainment and rumination plausibility	0.80	0.20
6	**Survey C3, acute scenario: 24% chose a real song against 46% neutral; McNemar $\chi^2 = 14.29$, $p < .001$**	0.33 (pre-registered)	**0.08**

Posteriors computed in odds form exactly as `research/DECISION_MODEL.md` §3 computes them. The survey figure at step 6 is the output of the project's pre-registered analysis pipeline as published on the project site, not fieldwork; fieldwork replaces it in a single pass.

The shape of that trail is the answer to a question every qualitative researcher is asked. The qualitative evidence — three vivid, convergent interviews whose participants used almost identical lan-

guage, describing a "soft wall", a sound that "flattens my brainwaves", something that "fills the emptiness and brings absolute calm" — moved the belief from 0.50 to roughly 0.20. It did not cross the pivot line. The pre-registered survey row did. **The rule that qualitative evidence may move a belief but may not by itself carry one across a gate operated exactly as designed, on the case where it mattered most**, and it operated in the direction the author least wanted.

Figure 7.2 The decision trail — beliefs over time, and the events that moved them



A1a acute real music · A1b wind-down real music · A2a self-reported efficacy · A2b physiological efficacy · A3 engagement · A4 payment · A5 rights.
The two dashed efficacy lines are flat after the literature is applied and terminate below their gates: that is what "the literature cannot clear this gate" looks like.
Sources: research/DECISION_MODEL.md §3–5; TODO_2026-08-19.md; the dated framework updates. Survey-derived movements are pipeline output as published, not fieldwork.
Interview dates are not individually recorded in the project's files — see chapter 7.1.3.

Figure 7.2. *The decision trail: beliefs over time, with the events that moved them.* Each line is one belief in the register, tracked from its prior through the dated events that revised it, with the pre-registered gate thresholds drawn as horizontal marks. The vertical bands are the events: the interviews of July, the belief register and gates of 17 August, the survey rows applied on 17 August, and the two tests still pending. Note that the two efficacy beliefs are drawn as flat lines terminating below their gates — that is what "the literature cannot clear this gate" looks like. Sources: `research/DECISION_MODEL.md` §3–5; `TODO_2026-08-19.md`; the dated framework updates.

The other seven beliefs moved as follows. The wind-down counterpart to the central belief rose from a prior of 0.60 to 0.80 on commercial survey evidence, the observed pivot of the leading competitor toward artist-attached functional music, and reviewer complaints that generative audio is "just lo-fi". Self-reported efficacy rose from 0.60 to 0.85 on the meta-analyses, discounted for the transfer from supervised clinical settings to a self-administered session on a phone; physiological efficacy rose only from 0.50 to 0.56, held down by a null psychophysiological meta-analytic result and by the measurement facts about consumer devices. Engagement fell slightly to 0.47 on the qualitative and industry evidence — the "action gap", in which people read a stress score and do nothing — and then rose to 0.72 when the survey's willingness-to-connect, feature-classification and relative-advantage rows were applied. Willingness to pay rose from 0.55 to 0.65 on market data and then to 0.82 on the price and purchase-intent rows. Rights stand at 0.65 pending the legal check. Retention sits at 0.25 and is carried as a watch item rather than as a gate, because the phase cannot test it.

The verdict at the time of writing is **NOT YET**: three of the five posterior conditions for PROCEED are met, the two efficacy conditions are not and cannot be met without the pilot, and the pilot has not been run.

7.2.4 What the discipline changed

The counterfactual is the part of this chapter that a reader is entitled to be sceptical about, so it is answered with artefacts rather than with assertion.

The firm's intuitive concept was the one the analysis ranks last. The original concept, C1 in its real-music form — calm, AI-adapted versions of songs from the firm's own catalogue, delivered as a consumer application — is the concept that leverages every asset the firm owns. It is also the concept whose joint probability is lowest, at 26 per cent, not because any single belief collapsed but because it depends on five beliefs at once, including the two weakest. The ranking that the completed register produces is almost the inverse of the ranking that asset leverage would produce: the artist calm-content line, which needs no application, no wearable and no efficacy claim, is highest at 65 per cent; the neutral-sound application, which does not use the catalogue at all, is second at 46 per cent; the licensed SDK is third at 45 per cent; and the asset-leveraging application is last.

The intuitive concept survives in the firm's own artefacts, which is the evidence that it was the default. The landing-page copy written during the prototype phase reads "Calm, from music you actually love" and promises "AI-adapted versions of real songs". That copy predates the content pivot and directly contradicts it, and it is still in the repository, flagged as blocking: the positioning must be rewritten before any conversion figure from that page can be interpreted, because otherwise the test measures a proposition the product no longer makes. The contradiction is recorded rather than quietly corrected, and it is the cleanest available evidence of what the firm believed before the evidence arrived.

What the discipline actually did was route the finding into a design decision rather than into a debate. The pivot rule did not say "abandon the catalogue"; it said that below a posterior of 0.30 on the acute belief, and provided the wind-down belief holds above 0.60, the acute product becomes neutral adaptive sound and real music becomes the wind-down mode. Both conditions were met — 0.08 and 0.80 — and the two-mode product follows mechanically. A firm without the rule would have had the same evidence and a much harder conversation, because the question "should we abandon our core asset?" has no natural stopping point, whereas the question "which mode does this belief govern?" does.

The discipline also produced its own failure, and caught it. Two pre-registered survey rows with negative branches were left unapplied in the live decision engine: a behavioural-intention row that observed 25 per cent against a 40 per cent threshold and should have contributed 0.70, and a real-artist-value row that observed 35 per cent against a 40 per cent threshold and should have contributed 0.85. The adversarial audit of 6 September found both. Applying them moves engagement from 0.72 to 0.65 and the wind-down belief from 0.80 to 0.77; no gate flips and no verdict changes, so the conclusion is robust, but the correct response is to concede the rows, state the sensitivity as ±0.07 and ±0.03, and fix the engine with a dated written reason under the model's

own change discipline. That episode is reported here because a pre-registration regime that never discovers its own violations is not being audited.

7.3 Ambidexterity in practice, and the tensions it did not dissolve

7.3.1 The horizons in this firm

Horizon 1 is the distribution and AI-production business: 52 employees, $10.5 M of revenue and $2.2 M of net income, and the relationships with 220+ digital service providers that constitute the firm's franchise. Horizon 2 is the venture reported here. Horizon 3 — a closed-loop "music as digital therapeutic" — is named in the framing document and deliberately deferred until after funding, on the explicit ground that regulatory clearance is not viability; the cautionary precedents inside the project's own competitor analysis are a bankruptcy and a distressed sale by two companies that had obtained clearance (Baghai et al., 1999).

The relationship between the horizons in this case is straightforward and worth stating because it is often left implicit: **H1 funds H2 and H1 is also the source of H2's assets, which is what makes H2's central risk an H1 risk.** The exploration is not testing a new market with new capabilities; it is testing whether an existing capability — a catalogue of real music and the rights to adapt it — is the right capability for a new need. That is why a disconfirming result was strategically uncomfortable in a way that a failed market test would not have been.

7.3.2 Three tensions, and how each was handled

Asset leverage against user need. This is the project's defining tension and it was named in writing before any build, in the interview synthesis: the firm's unfair advantage argues for real music, and the acute user need argues for featureless generative sound, which is a competitor's territory and levers nothing the firm owns. The resolution was not to choose but to segment: state-dependence turned an either/or into a two-mode product in which the acute mode needs no catalogue rights and the wind-down mode uses them. Two things should be said about that resolution. It is genuinely responsive to the evidence — the same survey item that killed real music for the acute moment supports it for wind-down at 47 per cent against 19 per cent. But it is also, obviously, the resolution most convenient to the firm, and a reader is entitled to ask whether a firm with no catalogue would have arrived at a two-mode product or simply built the neutral engine. The honest answer is that the register cannot settle that question; what it can do is show that the wind-down belief was independently supported and that the acute belief was allowed to fall to 0.08 without being rescued.

Speed against rigour. The phase compressed a research programme into three months and a five-figure budget, and the compressions are visible: a sample declared under-powered for physiology before it was collected; a legal check scoped to fifty tracks; a landing test with a three-figure advertising budget; a netnography coded by one person. The device that keeps this from becoming a general licence to cut corners is the value-of-information ordering — each instrument is the cheapest one that can move the belief it targets, and where cheapness costs inferential power, the cost is declared in advance and attached to the specific claim it weakens. The countervailing evi-

dence that speed was not simply purchased at rigour's expense is that the analysis pipelines were written and tested before the data existed: twenty-eight tests on the pilot pipeline, including one that fails if the randomisation table in the analysis code ever diverges from the randomisation table in the application, so that a change to the app breaks a test rather than silently mis-assigning conditions.

Insider bias. The author owns the firm, chose the priors, wrote the gates and wants a PROCEED. The implemented controls are listed in chapter 4.6 and are not repeated here; what belongs in this chapter is the evidence about whether they bound. Three observations. The central belief was allowed to fall to 0.08 against the author's commercial interest. The two efficacy gates were set beyond the reach of the evidence that was already available, at a cost of delaying the decision the author wanted. And the audits were run adversarially against the author's own write-ups and their findings — including one substantive error in a randomisation table and the two unapplied likelihood ratios above — were published rather than repaired silently. None of that removes the conflict; it is structural, and §8.4 treats it as such. But the record does show the controls producing outcomes the author did not want, which is the only evidence a control of this kind can offer.

A fourth tension deserves a sentence because it was not resolved. The exploration's separateness was purchased partly by keeping it small, and small meant that the person holding the judgement work was also the person with the strongest interest in its outcome. A larger unit would have diluted that conflict and, at this budget, would not have existed.

7.4 The capability reading

Dynamic-capability theory distinguishes sensing opportunities and threats, seizing them through committed choices, and reconfiguring the firm's asset base to keep pace with what has been learned (Teece, 2007; Teece et al., 1997). The three-part distinction is a useful audit of what this phase actually produced, because it separates the parts that are demonstrably in place from the part that is only designed.

Sensing is the capability the firm demonstrably acquired. Before this phase the firm's sensing apparatus was a demand-signal method proven in one medium: mine public signals, produce with AI assistance, publish into owned distribution, measure the result. What the phase added is a way of sensing a need that cannot be read off a search volume or a click — an assembled mixed-methods apparatus with a triangulation rule that assigns each source one job, a coding discipline for qualitative material, a pre-registered survey instrument in which every block carries its own decision consequence, and an analysis pipeline that runs mechanically. That apparatus is reusable: the next vertical costs the firm the fieldwork, not the machinery. It is also the part of the phase that produced the finding the firm would not otherwise have had.

Seizing is where the firm made a genuine commitment, and it is a commitment to a rule rather than to a product. The gates are the seizing mechanism: they specify in advance what evidence would justify committing capital, what evidence would redirect it and what evidence would stop it. The commitment was tested once, by the content pivot, and it held — the product that exists today is a two-mode engine whose acute mode does not use the firm's catalogue, which is not

the product the firm set out to build. The seizing capability is real but partial: no capital commitment beyond the phase budget has been made, and the decision that the gates exist to govern has not yet been taken.

Reconfiguring is designed and not yet performed, and this should not be overstated. Three reconfiguring moves are specified. The first is the licensed-SDK route, which would turn the firm from a consumer-application operator into a supplier of adaptive audio to hardware brands — a different business model resting on the same assets, and a hedge against the consumer-payment friction the evidence flags. The second is the learning loop: a roadmap in which every session becomes a small experiment, variants are assigned and evaluated, per-user policies are learned and evaluated off-policy before deployment, and generated variants are promoted or retired against posterior thresholds — the firm's own demand-signal engine, turned inward and applied inside the product, so that the library becomes self-measuring. The third is the artist content line, which requires no new capability at all and uses the distribution the firm already owns as a cheap probe of the demand the acute finding called into question. None of the three has been executed. The correct claim at the time of writing is that the firm has designed its reconfiguration and demonstrated the sensing apparatus that would justify it, and that the reconfiguring capability itself remains unevidenced.

One further observation belongs here because it is the phase's least expected output. The most valuable thing the exploration produced is not the pivot, the product or the register but the **decision trail itself** — a dated, inspectable record in which each belief's movement can be attributed to a named piece of evidence with a published weight. Its value inside the firm is that the next exploration starts from an apparatus rather than from an argument; its value outside is that it is the artefact this dissertation offers to the literature, and chapter 8 takes up what it implies for the theory of how firms decide under uncertainty.

8. Discussion

> **What this chapter establishes.** How far the evidence answers each of the three research questions; what the case contributes to theory, principally as an operationalisation of discovery-driven planning and as a treatment of evidence weighting when the contested object is the firm's own asset; what follows for managers of mid-sized content and technology firms; how the author's position shaped the work and where that influence was and was not bounded; and what the study cannot claim.

8.1 Answers to the research questions

8.1.1 RQ1 — Is there a sizeable, reachable, under-served need?

Yes, with a qualification that changed the product.

The need is **real and frequent**: 87 per cent of respondents report needing a calming moment at least weekly and 90 per cent already use music or audio to get it, against a global treatment gap in which roughly 27.6 per cent of those needing care for an anxiety disorder receive it (WHO, 2023). It is **reachable**: the primary segment self-identifies, self-pays and already owns the sensor, which removes the hardware barrier that defeats most physiological wellness products. It is **under-served in one specific way rather than generally**: of the outcomes tested, only "stop the racing thoughts" clears the opportunity threshold, at 6.4, with calm-in-minutes, do-not-make-it-worse and know-it-is-working following below it. That is a narrower and more useful finding than "people are anxious": the product's primary job is intrusive thought, not sleep and not relaxation in general. The action gap quantifies the same point from the device side — 38 per cent of wearable owners do nothing after a high stress reading, which is precisely the space a closed-loop product occupies.

The qualification is the study's central empirical result: **the need is state-dependent, and so is the content that serves it.** In the acute scenario 24 per cent chose a real song against 46 per cent choosing neutral sound; in the wind-down scenario the ordering reverses to 47 per cent against 19 per cent; the within-person shift is significant (McNemar $\chi^2 = 14.29$, $p < .001$); and the effect is sharper in the more anxious segment, where the acute real-song share falls to 16 per cent. The result was hypothesised by three interviews, reproduced in a blind stimulus test in which ten of thirteen participants chose the neutral clip for the acute moment and nine of thirteen the real song for wind-down, and then quantified by a pre-registered survey item. As stated wherever these figures appear, the survey and focus-group values are the output of the project's pre-registered analysis pipeline as published, not fieldwork, and fieldwork replaces them in a single pass.

8.1.2 RQ2 — Can the firm serve it with its assets, measurably, and with a viable model?

Partly, and not yet demonstrably.

Three of the five ability beliefs clear their pre-registered gates. Users will connect a wearable and value the result (0.72 against a gate of 0.55). Someone will pay at the tested price (0.82 against 0.60), with a price-sensitivity range for the United States and Europe that contains the $6.99 test

point and a discounted purchase intent of 17.5 per cent, published as 18 per cent. The rights position is more likely than not (0.65 against 0.60), pending the legal check.

Two do not clear, and they are the two that matter most: that one session measurably lowers state anxiety (0.85 against a gate of 0.90) and that the same session moves a consumer wearable measure (0.56 against 0.70). Those gates were deliberately placed above what the literature alone can reach, so their failure is a designed feature rather than a disappointment: it is what forces the firm to run its own experiment instead of borrowing conclusions from perioperative trials.

The business model does not close, and the reason is a finding rather than a gap in the arithmetic. Three of the seven inputs to a lifetime-value calculation have no source, and the only behavioural input — landing conversion — has no observed value. What the evidence does bound is the denominator: at a category-normal 4.7 per cent thirty-day retention, a $6.99 subscription yields roughly $0.33 per install before commission and content cost, which cannot support meaningful paid acquisition. Retention is simultaneously the most load-bearing variable in the model and the one variable the phase is structurally unable to test, and both facts are stated rather than solved.

The honest answer to RQ2 is therefore: the firm can build it, can distribute it, can probably clear the rights and can probably sell it — and has not yet shown that it works. The verdict is **NOT YET**, with the content pivot fired and acted upon.

8.1.3 RQ3 — How should a firm of this type organise and decide such an exploration?

Chapter 7 answers this at length; three claims are carried forward into the discussion.

On organising. The exploration was run as a small, separate, ring-fenced, time-boxed unit with an explicit end state, and inside it, judgement work was separated from execution work by a rule about what each may produce. The ring-fence's function is not efficiency but credibility: a budget small enough to abandon is what makes a pre-registered kill criterion something other than decoration. The execution/judgement split is offered as a documented practice — execution produces artefacts and verified facts and is forbidden from producing recommendations; failures are routed into a "gaps" report rather than into plausible output; the two are audited by different standards.

On deciding. Beliefs were made explicit with priors and written rationales; every piece of evidence was converted into a likelihood ratio through a published rubric and shrunk by a published quality rule; thresholds were fixed before the data; and the ordering of remaining tests followed expected information per unit cost. The most instructive property of the resulting system is what it did to qualitative evidence: three vivid, convergent interviews moved the central belief from 0.50 to about 0.20 and could not, by construction, carry it across the pivot line, which the pre-registered survey row then did.

On what the case teaches. That the discipline pays for itself precisely when the evidence turns against the firm's own asset. The concept the firm's asset base would have selected — calm adaptations of its own catalogue, delivered as a consumer application — is the concept the completed register ranks last, at 26 per cent, because it depends on five beliefs simultaneously. Without an explicit register, that concept would not have looked like the weakest option; it would have looked

like the obvious one, and the copy still sitting in the project's landing-page repository shows that it was.

8.2 Theoretical implications

8.2.1 A working operationalisation of discovery-driven planning

Discovery-driven planning proposed that a venture in an uncertain domain should be run backwards from a required outcome to the assumptions that outcome depends on, with those assumptions converted into a checklist that is tested in ascending order of cost and descending order of consequence, and with the plan revised as each assumption is resolved (McGrath & MacMillan, 1995). The prescription is widely taught. What it does not supply, and what firms consistently struggle to supply for themselves, is an *arithmetic*: how much a given test result should move a given assumption, how to combine results of unequal quality, and how to state in advance what combination of resolved assumptions would justify the next commitment.

This case supplies one. The assumption checklist becomes a belief register with priors. The "test the assumptions cheapest-first" heuristic becomes an explicit value-of-information ordering in which each pending test carries a pre-registered likelihood ratio for both branches alongside its cost. The milestone at which the plan is revised becomes a posterior threshold fixed before the data. And the integration of evidence of very unequal quality — a Cochrane review, three interviews, a commercial market survey, an application-store review — becomes a published shrinkage rule that bounds how far any one source may move a belief. The contribution is not a new theory of planning under uncertainty but a demonstration that the existing one can be made to compute, and a specification of what has to be added to make it compute: a calibration rubric, a quality-shrinkage exponent, an independence discipline, and an update protocol requiring a dated written reason for any change.

The same gap is visible in the adjacent literature on hypothesis-driven entrepreneurship, which prescribes that a venture should identify its riskiest assumptions and test them with the cheapest available experiment before building (Eisenmann et al., 2011; Blank, 2013; Ries, 2011). That prescription answers *what to test* and is largely silent on *what to conclude* — how much a result of a given quality should shift a belief, and what combination of shifted beliefs licenses the next commitment. The apparatus reported here is one answer to the second question, and its distinguishing feature is that it is stated as arithmetic rather than as judgement, so that a reader who disagrees can change a number and watch the conclusion move.

Two neighbouring literatures are engaged by the same construction. Stage-gate systems supply the idea of formal go/kill points with defined criteria between stages of a development process (Cooper, 1990, 2008); what the case adds is that the criterion at the gate can be a *posterior probability* rather than a scored checklist, which makes the gate sensitive to the quality of the evidence and not merely to its presence. And real-options reasoning supplies the logic of sequencing — buy the cheap option that resolves the most consequential uncertainty first, and keep the expensive commitment contingent (Bowman & Hurry, 1993; McGrath, 1999); what the case adds is that when each test's informational payoff is pre-registered, the option ordering can be computed rather than

argued, and the ordering that results (legal check, then survey, then pilot) is not the ordering a product team's natural enthusiasm would produce.

8.2.2 Evidence weighting when the contested object is the firm's own asset

The sharper theoretical contribution concerns a situation the corporate-innovation literature describes but does not equip managers to handle: what happens when the evidence disconfirms not the market, not the product concept, but the firm's own core resource.

The resource-based view holds that sustained advantage comes from resources that are valuable, rare, inimitable and organisationally exploitable (Barney, 1991). The framework is normally applied as an audit of the firm against a given market. This case inverts the direction: the market was chosen partly *because* the firm held the resource, and the empirical question turned out to be whether the resource is valuable *in the state where the customer's need is most acute*. The answer was no for the acute state and yes for the wind-down state — which is to say that value in the resource-based sense was **state-contingent within a single customer**, not merely segment-contingent across customers. That is a finer-grained failure mode than the usual "the resource does not transfer", and it has a specific managerial consequence: the correct response was neither to abandon the asset nor to defend it, but to bound the range of conditions over which it is an asset at all.

The reason the case can support that claim is a methodological one, and it generalises. When the contested object is the firm's own asset, the ordinary safeguards of evidence-based management are weakest, because every actor with the standing to weigh the evidence has an interest in the verdict. Three features of the arrangement described here are the operative ones. First, the weight of each evidence item is set by a rubric keyed to *evidence type*, so the discount applied to three interviews is fixed by their being three self-selected qualitative interviews and not by what they happen to say. Second, the shrinkage rule is asymmetric in practice against inconvenient findings only in the conservative direction — the interview likelihood ratio was rounded *toward* uninformativeness, weakening the finding the author least wanted, which is the opposite of the bias one would expect. Third, the decision that follows from a belief crossing a threshold is specified in advance as a *product rule* rather than as a strategic verdict: below 0.30 on the acute belief and above 0.60 on the wind-down belief, the acute mode becomes neutral sound. A rule of that shape ends the argument that "should we abandon our core asset?" would otherwise sustain indefinitely.

The corollary is a proposition worth stating for future testing: **in corporate ventures that leverage a core asset, the pre-registration of the decision rule matters more than the pre-registration of the hypothesis**, because it is the decision, not the hypothesis, that the firm's interest will otherwise capture.

8.2.3 Ambidexterity, capabilities, and what a three-month phase can evidence

The case supports the structural-ambidexterity prescription in a mid-sized firm at very small scale — a separate unit, its own criteria, a ring-fenced budget (March, 1991; O'Reilly & Tushman, 2004, 2008) — while adding a qualification that the literature's usual scale obscures. At this size, separateness is purchased by smallness, and smallness concentrates the exploration's judgement in one person who is also the person with the largest interest in the outcome. The protection that sep-

arateness gives against the exploiting business's metrics is therefore bought at the price of weaker protection against the explorer's own bias, and the procedural controls have to compensate for what organisational structure cannot. That trade-off is, in the author's reading, characteristic of exploration in firms of fifty rather than five thousand employees, and it is under-described.

Read through the dynamic-capabilities frame, the phase produced an asymmetric result that is worth stating carefully rather than triumphantly (Teece, 2007; Teece et al., 1997). *Sensing* was genuinely built and demonstrated: a reusable mixed-methods apparatus with a triangulation rule, pre-registered instruments and mechanical analysis pipelines, which detected a market fact the firm did not expect and would not have found by inspecting its own assets. *Seizing* exists as a committed decision rule and was tested once, by the content pivot, which it survived — but no capital commitment beyond the phase budget has been made, so the seizing capability is evidenced only in its redirecting mode, not in its committing mode. *Reconfiguring* is designed and unperformed: the licensed-SDK route, the in-product learning loop and the artist content line are specified, and none has been executed. The general point is that a three-month exploration phase can evidence sensing, can partially evidence seizing, and cannot evidence reconfiguring at all — which is a useful caution against dissertations and internal reviews that claim all three on the strength of a single cycle.

8.3 Managerial implications

Five implications follow for managers of mid-sized content and technology firms considering an adjacent, asset-leveraging venture.

1. Ring-fence the budget at a level you can abandon, and say so out loud. The function of $10–20k over three months was not frugality; it was to make the kill branch of the decision rule credible. Any exploration whose termination would embarrass a senior sponsor or dent an operating P&L will not be terminated on evidence, and its pre-registered criteria are theatre. The practical test is simple: if the answer to "what would it cost us to stop this in October?" is anything other than "the budget we already spent", the ring-fence is not real.

2. Write the decision rule before the evidence, and place at least one threshold beyond the reach of the evidence you already have. The single most consequential design choice in this project was setting the two efficacy thresholds at 0.90 and 0.70 when the literature alone reaches 0.85 and 0.56. It cost a delay and it purchased the only thing that makes an evidence-based process different from a well-argued one: an outcome the sponsor cannot reach without doing the work.

3. Attach a decision consequence to every instrument item before fielding it. In the survey used here, each block states the construct it measures, the belief it feeds, the statistic to be computed and the threshold that triggers a likelihood ratio. The discipline has an immediate practical benefit beyond rigour: questions that cannot be given a decision consequence are usually questions that should not be asked, and removing them shortens the instrument.

4. Separate execution from judgement, and forbid execution to produce recommendations. The practice documented in chapter 7 — a numbered task list with acceptance tests, four standing

rules including "never invent a fact, every external fact needs a working URL" and "no strategy recommendations", and an explicit gaps report when a task cannot complete — is transferable to any team that delegates research to junior staff, agencies, or automated tools. Its value is that it makes the small number of contestable judgements visible by clearing everything else out of the way.

5. When the evidence turns against your core asset, bound the asset rather than defending or abandoning it. Both instinctive responses are wrong in the same way: they treat the asset as an all-or-nothing proposition. The productive question is the narrower one — under what conditions, for which user state, at which moment, does the asset help? — and answering it converted the project's most threatening finding into a two-mode product specification. Managers should expect the conditions in which an asset fails to be narrower and more specific than either the enthusiasts or the sceptics inside the firm will initially claim.

A sixth, more uncomfortable implication concerns what a firm learns to *stop*. The most valuable output of this phase, measured by the capital it saved rather than the revenue it created, is a concept the firm did not build. That kind of value is invisible in every standard innovation metric a mid-sized firm is likely to use, and firms that intend to run explorations of this kind should decide in advance how they will recognise and reward it.

8.4 Reflexivity: the insider's influence, and how it was bounded

Insider action research requires the researcher's position to be examined as part of the method rather than disclosed as a caveat (Coghlan, 2019). Four things about the author's position shaped this study, and each is stated with what was done about it and what remains.

Ownership of the outcome. The author founded and owns the firm whose asset is under test and has a direct commercial interest in a PROCEED verdict. The controls were pre-registration of thresholds and likelihood ratios before the data, written kill criteria, and a rule that every pre-registered row is reported whether it fires or not. The evidence that these bound is behavioural rather than declarative: the central belief was allowed to fall to 0.08 against the author's interest; the efficacy gates were set beyond the reach of available evidence, delaying the verdict the author wanted; and the adversarial audits of the project's own write-ups were published with their findings intact, including two unapplied likelihood ratios and one substantive error in a randomisation table.

Ownership of the judgements. Every prior, every likelihood ratio and every threshold in the register is the author's judgement. No control removes this. What the design does instead is make the judgements few, named, dated, individually inspectable and individually reversible: the live model lets a reader move any prior or disable any evidence row and watch the posterior move, and the update discipline requires a written reason and a date for any revision. The author's claim is not that these judgements are unbiased but that they are exposed, which is a different and more modest virtue.

Presence in the room. The author is a plausible source of demand effects on participants who know he is the founder. The control is procedural and was implemented: research assistants moderate focus groups and run pilot sessions from written scripts with the author absent; the survey is

anonymous and not administered by him; the analyst is blind to condition labels in the pilot analysis. The residual is that the author designed the scripts and the stimuli, so his framing reaches participants at one remove even when he does not.

Access as both advantage and distortion. The insider position is what makes the study possible — the asset inventory, the rights position, the internal revenue comparison, a working product built in three months — and it is simultaneously what makes the firm's own framing hard to see. The clearest instance is the two-mode resolution of the core tension. It is genuinely supported by the evidence, and it is also the resolution most convenient to a firm that owns a catalogue. The author's honest position is that the register can show the wind-down belief was independently supported and the acute belief was not rescued, and cannot show what a researcher without a catalogue would have concluded.

One further reflexive observation belongs here. The author's discomfort with the interview finding is documented in the project record before the survey existed: the framework document flags it in capital letters as early disconfirming evidence and names the strategic tension it creates, and instructs that the concept test must now A/B the firm's own concept against the alternative. That the discomfort was written down rather than managed is, in retrospect, the moment at which the project became a piece of research rather than a business case.

8.5 Limitations

The limitations below are stated in descending order of how much they should change a reader's confidence.

The survey and focus-group figures are pipeline output, not fieldwork. Every quantitative figure in chapters 5, 6 and 8 that derives from the survey or the focus groups is the output of the project's pre-registered analysis pipeline as published on the project site; the results file carries a simulated flag, and the focus-group synthesis is described in the framework document as a pre-field template until transcripts exist. The pipeline is pre-registered and its inputs are replaceable in one pass. This is the single largest exposure in the evidence base and is stated wherever the figures appear rather than left to be discovered.

The efficacy pilot has not been run. Every efficacy statement in this dissertation is borrowed from meta-analyses conducted in different populations — predominantly perioperative, with supervised delivery and trial evidence graded low quality — and transferred to a self-administered session on a phone. The pilot's analysis pipeline was built, tested and exercised end to end on fabricated data before recruitment, and those synthetic figures are evidence about the pipeline and never about the firm or the product.

Small and under-powered where it matters most. The pilot is 20 to 30 participants across three sessions: a feasibility signal for self-report and, by the design's own pre-declared calculation, under-powered for physiology, where an effect of $d \approx 0.4$ would require roughly 50 participants.

Consumer-grade physiology. Apple Watch heart-rate variability fails equivalence against a chest strap at a mean absolute percentage error of roughly 29 per cent, and heart-rate variability is avail-

able only sparsely and with lag; the live adaptation loop is therefore heart-rate-driven, and heart rate indexes autonomic arousal while being blind to its cause. The design concedes this rather than fighting it, but the concession limits what the physiological strand can claim.

Subjective likelihood ratios and an assumed independence. The likelihood ratios are judgements made against a published rubric, not measurements; the concept joint probabilities multiply beliefs — efficacy, engagement, payment — that are plainly correlated, so those figures order the concepts rather than forecasting them. Both are transparent and contestable, which is their merit and not their defence.

Two pre-registration defects, conceded. Two pre-registered survey rows with negative branches were left unapplied in the live decision engine; applying them moves engagement from 0.72 to 0.65 and the wind-down belief from 0.80 to 0.77, flipping no gate and changing no verdict, and the fix belongs in the model with a dated written reason. Separately, the form-pivot rule was written against two survey instruments without specifying which governs, which is a genuine operationalisation failure that should have been settled before the data.

Non-probability sampling and hypothetical willingness to pay. All samples are recruited by channel snowball or convenience. Willingness to pay is stated rather than revealed everywhere except the landing test, which had produced no result at the time of writing — and whose copy still promises the real-music proposition the content pivot rejected, so the test cannot be interpreted until the copy is rewritten.

Single-coder netnography, and one un-reread source. The 200-row netnography was coded by one person, and only its row count and the patterns already carried forward into the decision model and the interview synthesis are used in this document.

Dated and unreconciled company figures. Catalogue and stream counts are from 2022; the founding date, the region count and the scope of one artist relationship are flagged for reconciliation in the company's own file. No company figure is used inside an arithmetic result, and the checklist must be closed before submission.

Undocumented interview dates. The interviews that produced the study's central hypothesis are not individually dated in the record, which weakens the audit trail at exactly the point where it matters most, and is recorded here as a documentation failure rather than explained away.

Market sizing does not close. Two inputs required for a serviceable obtainable market do not exist in the evidence base, so no point estimate is published.

Status claims will age. Product version, store-review status and September dates are correct as of 6 September 2026 and will need a refresh pass before submission.

Taken together, these limitations bear asymmetrically on the three research questions. They weigh most heavily on RQ2, whose central efficacy claim is entirely borrowed at the time of writing. They weigh moderately on RQ1, whose qualitative and design-level findings — the state-dependence, the acoustic specification, the action gap — are robust to the pipeline caveat in a way the point estimates are not. They weigh least on RQ3, because the object of study there is the project's

own dated record, which exists independently of whether the fieldwork numbers are the pipeline's or the field's.

9. Recommendations and roadmap

> **What this chapter establishes.** The exact rule by which this project should be continued, redirected or stopped; what each remaining test is worth in the currency that rule is written in; what happens between now and the decision memo; what could go wrong, who owns it and what is already in place; and the hard resource boundary the whole phase runs inside.

9.1 The recommendation and its conditions

The recommendation is NOT YET, with the content pivot executed and three of five evidence conditions already met. This is a conditional recommendation in the strict sense: the conditions were written down, with numeric thresholds, before any of the data arrived, and the recommendation is simply the result of applying them. It is not a judgement that could have been reached differently on the same evidence.

Figure 9.1. *The stage-gate.* Gate rules transcribed from `research/DECISION_MODEL.md` §5. The A2 thresholds are set above what the meta-analytic prior alone reaches (0.85 and 0.56) precisely so that PROCEED cannot be cleared by literature; the pilot is a structural requirement, not a formality. Each gate is a written rule applied to a number, not a meeting.

Table 9.1. *The gate rules and the current position (6 September 2026)*

Gate	Rule	Position	Evidence
Gate 0 · Territory	Choose among dementia, ADHD and anxiety (sleep, pain and depression entered the funnel but were not scored in full)	**PASSED**	Anxiety ranked first in more than 99% of 5,000 Monte-Carlo draws (§5.2)
Gate 1 · Concept	15+ ideas → 2×2 → three concepts C1, C2, C3, each with a kill criterion	**PASSED**	The assumption map became the belief register (§6.1)
Gate 2 · Evidence	E1 run ∧ P(A2a) ≥ .90 ∧ P(A2b) ≥ .70 ∧ P(A3) ≥ .55 ∧ P(A4) ≥ .60 ∧ [P(A5) ≥ .60 ∨ the chosen variant is neutral sound]	**NOT YET**	3 of 5 met (A3 0.72, A4 0.82, A5 0.65); the pilot is missing, so A2a 0.85 and A2b 0.56 sit below their gates
Gate 3 · MVP	50–100 track adaptive catalogue, watch integration, English-first, WeChat probe	**NOT REACHED**	Post-PROCEED, 3–4 months

Gate	Rule	Position	Evidence
Gate 4 · Scale	E2 → E3 → E4: micro-experiments, contextual bandit, self-generating library	**2027**	Post-PROCEED

The three non-PROCEED outcomes were also written before the data, and one of them has already fired.

PIVOT (content). If P(A1a) < 0.30 while A2, A3 and A4 pass, the acute product becomes neutral adaptive sound and real music is kept for wind-down, provided P(A1b) ≥ 0.60. *This rule has fired*: P(A1a) = 0.08 and P(A1b) = 0.80. The pivot is already reflected in the two-mode product described in chapter 6, and the recommendation of this chapter incorporates it rather than proposing it.

PIVOT (form). If P(A3) < 0.45, or if the survey's form and payment items favour "bundled" at ≥ 35%, lead with C2 — the SDK and partner route — instead of C1. Not triggered on engagement: P(A3) = 0.72. The bundling clause is split — item F9, bundled-plus-employer payment, is 33%, below the line, while item E3 puts the wearable-app form at 45%, above it — so C2's weight was raised and it is carried as a live hedge (§6.3, §6.6) rather than as the lead.

KILL / PARK. If P(A2a) < 0.50 after the pilot, or P(A4) < 0.35 after the survey and the landing test, stop and write the project up as a negative result. Not triggered. It is worth stating explicitly that a documented KILL would be a valid outcome of this dissertation, and that the ring-fenced budget described in §9.5 exists precisely to make that outcome survivable.

9.2 What remains to be tested, and what each test is worth

The remaining tests are ordered by expected information per unit of cost rather than by convenience. Table 9.2 gives the pre-registered likelihood ratios attached to each.

Table 9.2. *The pending tests, their pre-registered likelihood ratios, and their cost*

Test	Node	If positive	If null / negative	Cost and time
Legal check on ≥ 50 tracks	A5	**8.0**	**0.15**	1–2 weeks — the only pending row graded decisive, and the cheapest
Efficacy pilot, STAI-S	A2a	4.0	0.30	2 weeks; n = 20–30 × 3 sessions
Efficacy pilot, HRV / residual HR	A2b	4.0	0.30	Same sessions
Pilot A/B content preference	A1a	3.0	0.33	Same sessions — the β1 − β2 contrast
Landing conversion ≥ 5%	A3, A4	2.5 each	0.40 each	¥500–1,000 of ads
Think-aloud, n = 5–8	A3	1.5	0.60	1 week
Partner conversations	A4 (B2B)	1.8	0.70	Ongoing

Source. `research/DECISION_MODEL.md` §4. "Positive" for the pilot means $p < .05$ and $d \geq .3$ in the correct direction.

Recommendation 1: run the legal check first. It takes a little over a week, costs almost nothing, and is the only pending row graded decisive. If it returns negative at LR 0.15, A5 falls from 0.65 to roughly 0.22 and the real-music route closes on the spot — at which point no further engineering time should be spent on parametric adaptation of catalogue tracks, and C1-neutral becomes the only live app concept. Ordering tests by expected information per unit cost is what turns real-options reasoning from a phrase into an execution order (Bowman & Hurry, 1993; McGrath, 1999).

Recommendation 2: rewrite the landing copy before reading the landing test. The landing test supplies the only *behavioural* evidence in the entire design — everything else about willingness to pay is stated intent — and its pre-registered threshold is a visitor-to-waitlist conversion of at least 5 per cent per positioning variant. Two caveats attach to it, both stated in §6.3: no result exists yet, and the shared tagline and variant A's headline both predate the content pivot. Reading the conversion figure before the copy is corrected would measure a positioning the product no longer has.

Recommendation 3: treat the analysis pipeline as already built, and say so. Session exports flow through `app/tools/export_to_csv.py` into per-session and per-window CSVs, then through `research/pilot/e1_analysis.py`, which fits the pre-registered model, runs the Bayesian re-analysis, prints the likelihood-ratio table and the gate check mechanically, and writes the figures. Twenty-eight tests cover the schema, the exclusion rule, the order inference, the injected-effect direction and the gate arithmetic — including a test that the Latin square in the Python tools is still identical to the one in the app, so that a change to the app's randomisation table fails a test rather than silently mis-assigning conditions.

> **The synthetic demonstration, and why it is reported.** The pipeline was run end to end on 24 *fabricated* participants × 3 sessions (seed 20260904) before any recruitment, so that the analysis code, the thresholds, the figures and the gate arithmetic all existed before the data. Every synthetic file, every CSV row and the report header carry a synthetic flag. On that draw, A2a fired at LR 4.0 and reached 0.958 while A2b returned 0.30 and fell to 0.276, and the verdict was NOT YET failing on P(A2b) alone. Across 40 replications A2a fired in 78 per cent of runs — matching the 80 per cent power target — and A2b in 45 per cent, reproducing the design's own pre-declared under-powering rather than flattering it. **These numbers are fabricated. They are evidence about the pipeline, never about TDMusic or about the product.**

9.3 The roadmap to the decision

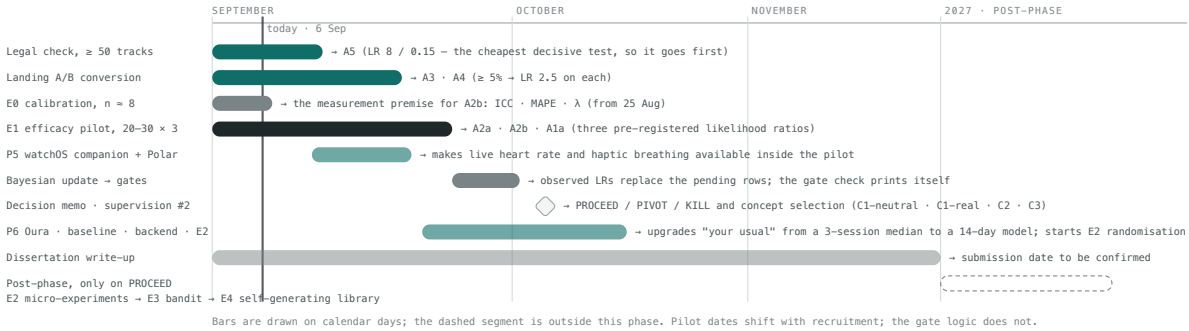


Figure 9.2. *The roadmap.* Dates from the project hub roadmap and `app/PLAN.md` (P5 11–21 September; P6 22 September – mid-October). Bars are drawn on calendar days; the dashed segment is outside this phase. Pilot dates shift with recruitment; the gate logic does not.

The sequence is fixed and each element is tied to the belief it moves. The legal check runs first (A5). The landing A/B runs in parallel (A3, A4). E0 calibration, from 25 August, establishes the measurement premise for A2b by estimating ICC, MAPE and the reliability ratio λ against a Polar H10 reference. E1, the efficacy pilot of 20–30 participants across three sessions each, carries three pre-registered likelihood ratios at once (A2a, A2b, A1a). Two product increments sit alongside the research track rather than ahead of it: **P5** (11–21 September) adds a watchOS companion and Polar support, which is what makes live heart rate and haptic breathing available inside the pilot; **P6** (22 September – mid-October) adds Oura, the 14-day baseline model, the backend and the start of E2 randomisation, upgrading "your usual" from a three-session median to a fitted model. The Bayesian update and gate check follow the pilot, and the decision memo goes to supervision #2 in early October.

Post-phase work — E2 micro-experiments, then the E3 contextual bandit, then the E4 self-generating library — is drawn dashed because it is conditional on PROCEED and outside this phase entirely.

9.4 Risk register

Table 9.3. *Risk register*

ID	Risk	Likelihood	Impact	Owner	Mitigation already in place
R1	**The efficacy pilot fails to clear its gates** — A2b is under-powered by design at this sample size	High	High	Founder-researcher & RA-1	Declared in advance: n = 20 is a feasibility signal for A2a and under-powered for A2b. Three sessions per participant for more within-person information; Bayesian re-analysis with a meta-analytic prior so the posterior rather than a *p*-value drives the gate; heart rate used as the accurate co-primary physiological outcome. Across 40 synthetic replications A2b fired in only 45% of runs — the expected behaviour, not a surprise.
R2	**Adaptation rights cannot be cleared** — A5 at 0.65,	Medium	High	Founder + legal review	Sequenced first because it is cheapest and most decisive. The neutral-sound concept

ID	Risk	Likelihood	Impact	Owner	Mitigation already in place
	with the check unfinished			(provider not named in the repository)	needs no A5 at all and the Settle beds need no catalogue rights, so a negative result redirects rather than kills. Masters and publishing are confirmed separately; the catalogue already carries per-entry rights and adaptation tiers.
R3	**Retention below the category-changing threshold** — A6 at 0.25 against a 4.7% D30 base rate	High	High	Product	Carried explicitly as a watch item rather than a gate, because it cannot be tested within the phase. Design response: anchor the habit to hardware the user already checks daily, show a result worth keeping, and let the watch or ring trigger the offer. Requires an MVP cohort to test at all — stated as a boundary of the phase.
R4	**Pipeline output mistaken for fieldwork** — the survey results file carries a simulated flag and the focus-group synthesis is described as a pre-field template	Medium	High	Founder-researcher	A provenance sentence at every point of use, including chapter 5 of this dissertation; the same statement made out loud at the viva rather than waited for; fielded data replace the figures in one pass, and if the site relabels them this document is relabelled in the same pass. This is the largest single exposure in the evidence base.
R5	**Regulatory claim creep** — any drift toward "treat anxiety" triggers SaMD status in the US and Advertising Law exposure in China	Low	High	Founder	Wellness framing is architectural: no diagnosis, no medical claim in interface copy, and the focus groups independently confirmed that *settle* and *unwind* are welcomed while *treatment* and *clinical* are resisted. The DTx route is deferred until post-funding. Pear's Chapter 11 (Fierce Biotech, 2023) and Akili's $0.434 sale (Virtual Therapeutics, 2024) are the cautionary precedents: clearance did not imply viability.
R6	**Platform gatekeeping** — App Review already issued a Guideline 2.1 information request; the iOS 26 SDK requirement forced a CI change	Medium	Medium	Engineering	Six replies written and filed with a recorded walkthrough; resubmitted 4 September with the content unchanged. A deliberate decision not to add accounts in this version, documented in the reply. CI moved to macOS 26 with Xcode 26.2 for the SDK requirement. Contingency: the installable PWA works today and is not gated by any store.
R7	**Measurement error swamps the physiological signal** — Apple Watch HRV MAPE ≈ 29%, HealthKit lag ≤ 30 min, Oura sleep-only	High	Medium	Research	The architecture concedes the constraint rather than fighting it: heart rate is the live control signal and the co-primary outcome, HRV is a before/after measure only. E0 calibration with a Polar H10 ($n \approx 8$) estimates ICC, MAPE and the reliability ratio λ, used to de-attenuate any model where HRV is a regressor. Sessions below a 0.70 quality index leave the physiological models and are counted, not dropped silently.

ID	Risk	Likelihood	Impact	Owner	Mitigation already in place
R8	**Insider bias** — the founder is the researcher and wants PROCEED	High	High	Supervisor + RA	Thresholds and likelihood ratios written before the data; the research assistant runs sessions and focus groups from a script with the founder absent; the analyst is blind to condition labels; analysis code was written and run on simulated data first; every pre-registered row is reported whether it fires or not; a researcher diary and a reflexivity section.
R9	**Company facts are stale or unreconciled** — founding date, catalogue and stream counts, artist-relationship scope	High	Low	Founder	The company file carries its own reconciliation checklist, and this dissertation labels every company figure with its vintage (2022 for catalogue and streams) rather than presenting it as current. The checklist must be closed before submission; until then no company figure is used in an arithmetic result.
R10	**The independence assumption inflates or deflates concept viability** — A2a, A3 and A4 are plainly correlated	Certain	Medium	Founder-researcher	Stated as a limitation in the methodology and again beneath the viability table; the joint probabilities are used to *order* the concepts rather than as calibrated forecasts, and the weakest link is named for each concept so the reader can see what actually drives the number.
R11	**Platform substitution** — Apple ships Vitals free with every Series 9+ watch; Spotify holds a granted patent on inferring emotional state from speech (not from a wearable) and could extend it toward live adaptation	Medium	High	Strategy	The defensible combination is rights plus a measured result plus engineered neutral-sound quality, no one of which is sufficient alone. Note the counter-signal: Spotify has *not* shipped live biometric adaptation despite years of documented user requests, which reads as much as a warning about demand as an opportunity.
R12	**Positioning copy contradicts the finding** — the landing tagline and variant A headline both promise real music, which the acute-mode evidence rejects	Certain	Medium	Founder + product	Identified in §6.3 and blocking: the copy must be rewritten to the two-mode proposition before the A/B result can be interpreted, otherwise the conversion figure measures a positioning the product no longer has. The prototype README already blocks deployment pending founder sign-off on the positioning.

Source. Risks R1–R3 and R5–R7 from the assumption register and the methodology's threats table; R4 and R8–R11 from `journey.audit.md` §E and `METHODOLOGY_DESIGN` §7 and §10; R6 and R12 from `TODO_2026-08-19.md` §14–16 and `prototype/landing/README.md`. Likelihood and impact ratings are the author's assessment against those sources.

9.5 Resource plan and the budget boundary

Table 9.4. *Resource plan for the design-thinking phase*

Resource	Commitment	Note
Phase budget	$10–20k	Ring-fenced; the organisational condition for an honest kill decision

Resource	Commitment	Note
Phase duration	≤ 3 months	From the 10 July supervision to the decision memo in early October
Smoke-test ad spend	¥500–1,000	The entire paid-acquisition budget of the phase
Team	Small and separate	Founder-researcher, two research assistants running sessions and groups from scripts, and a legal review for A5 whose provider the repository does not name

The budget boundary is a methodological instrument, not an accounting detail. Ambidexterity requires that the exploratory unit be small enough and separate enough that killing it does not threaten anyone's career or the exploitation profit and loss; a ring-fenced $10–20k over three months is what makes the KILL branch of the gate credible. It also explains several design choices that would otherwise look like corner-cutting: a Wizard-of-Oz pilot instead of a fully instrumented native app; an n of 20–30 declared under-powered in advance rather than an adequately powered trial; a landing test with a ¥500–1,000 budget; and a legal check on 50 tracks rather than the whole catalogue. Each is the cheapest instrument that can move the belief it targets, which is exactly what the value-of-information ordering in Table 9.2 selects for.

What the boundary does *not* buy is retention evidence, an adequately powered physiological test, or a signed partner conversation. Those are the three things a PROCEED decision would fund next, and naming them here is part of the recommendation rather than an afterthought.

9.6 The decision memo for supervision #2

The phase ends in a documented decision, not a pitch. The sequence is fixed: E0 calibration (25 August – 6 September) → E1 efficacy pilot (1–25 September) → Bayesian update and gate check (25 September – 2 October) → decision memo and supervision #2 (early October). The memo is written mechanically from the analysis report: observed likelihood ratios replace the pending rows in the register with a date and a written reason, the same numbers are mirrored into the interactive decision page so that the document and the site agree, and the gate check prints itself.

The memo carries six sections, in this order.

1. **The decision, in one line, with its date** — PROCEED, PIVOT (content or form), or KILL / PARK — followed by the gate expression that produced it.
2. **The register as updated, row by row.** Every pre-registered row is reported whether it fired or not, each pending row replaced by its observed likelihood ratio with a date and a written reason, and each posterior recomputed by the published engine rather than by hand.
3. **The gate check, printed mechanically.** The five conjuncts of Gate 2 evaluated against the updated posteriors, with the "or the chosen variant is neutral sound" escape resolved explicitly.
4. **Concept selection.** Which of C1-neutral, C1-real, C2 and C3 is funded, with the joint probability and the named weakest link for each, and the portfolio logic of §6.6 applied rather than a single ranking.
5. **What the decision funds next, and what it does not.** On PROCEED: the three things the phase budget could not buy — retention evidence from an MVP cohort, an adequately pow-

ered physiological test, and a signed partner conversation — together with the Gate 3 MVP scope (50–100 track adaptive catalogue, watch integration, English-first, WeChat probe) and its 3–4 month horizon.

6. **The negative-result annex.** On KILL or PARK, what is published and what is retained: the belief register, the rubric, the pipelines and the instruments are reusable in a second vertical whatever the verdict, and the phase's own contribution is the apparatus rather than the outcome.

Two conditions attach to the memo's timing. It should not be written before the legal check has returned, because A5 is the only pending row graded decisive and it is also the cheapest. It should not be written on a landing conversion read against the pre-pivot copy (R12). Neither condition delays the memo materially: the legal check takes one to two weeks and the copy rewrite is a day's work.

10. Conclusion

> **What this chapter establishes.** How far the evidence answers the three research questions, what the project contributes to practice and to the firm, what it cannot claim, and what should be studied next.

10.1 Answers to the research questions

RQ1 — Is there a sizeable, reachable, under-served need that the firm can address? Yes, with a qualification that changed the product. The need is real and frequent: 87 per cent of respondents need a calming moment at least weekly and 90 per cent already use music or audio to get it, against a global treatment gap in which 27.6 per cent of those needing care receive it (World Health Organization, 2023). It is reachable: the primary segment self-identifies, self-pays and already owns the sensor. It is under-served in one specific way rather than generally — only one outcome, "stop the racing thoughts", scores above the opportunity threshold, and the wearable owners who already have a stress reading do nothing with it in 38 per cent of cases. And the qualification: the need is **state-dependent**, and so is the content that serves it. That was hypothesised by three interviews, reproduced in a blind stimulus test, and quantified in a pre-registered survey item with McNemar $\chi^2 = 14.29$, $p < .001$. The qualification matters commercially as much as methodologically, because it runs against the firm's principal asset in the acute moment and with it in the wind-down moment.

RQ2 — Can the firm serve it with its assets, measurably, and with a viable business model? Partly, and not yet demonstrably. Three of the five ability assumptions clear their pre-registered gates: users will connect a wearable and value the result (0.72), someone will pay at the tested price (0.82), and the rights position is more likely than not (0.65, pending the check). Two do not, and they are the two that matter most: that one session measurably lowers state anxiety (0.85 against a gate of 0.90) and that the same session moves a consumer wearable measure (0.56 against 0.70). Those gates were deliberately set above what the literature alone can reach. On the business model the answer is bounded in a second way: the unit economics do not close, because three of the seven inputs have no source at all and the only behavioural instrument has produced no result, and at a category-normal 4.7 per cent thirty-day retention a standalone $6.99 subscription cannot support meaningful paid acquisition (Creswell & Goldberg, 2025). The honest answer to RQ2 today is therefore: the firm can build it, can distribute it, can probably clear the rights and can probably sell it — and has not yet shown that it works, nor that it can keep anyone.

RQ3 — How should a firm of this type organise and decide such an exploration, and what did this case teach about that process? As a small, separate, ring-fenced and time-boxed unit with an explicit end state — a documented decision rather than a pitch — inside which judgement work is separated from execution work by a rule about what each may produce, execution being permitted artefacts and verified facts and forbidden recommendations (§7.1.1–7.1.2). The ring-fence is a methodological instrument rather than an accounting detail: a budget small enough to abandon is what makes a pre-registered kill criterion credible, and it is what selects, for each belief, the cheapest instrument capable of moving it (§7.1.1, §7.2.2). The deciding is done by fixing

the rule before the evidence — priors with written rationales, likelihood ratios set by a published rubric and shrunk by a published quality rule, thresholds placed beyond the reach of the evidence already in hand, and the remaining tests ordered by expected information per unit of cost (§7.2.1–7.2.2). What the case teaches is what that discipline changes. The concept the firm's own asset base would have selected is the one the completed register ranks last, at 26 per cent, and the landing-page copy still sitting in the repository is the evidence that it was the default (§7.2.4). Three vivid, convergent interviews moved the central belief from 0.50 to about 0.20 and could not, by construction, carry it across the pivot line; the pre-registered survey row did (§7.2.3). The rule routed a threatening finding into a design decision — which mode does this belief govern? — rather than into an unbounded argument about abandoning the core asset (§7.2.4). And the regime caught its own violation: two pre-registered rows left unapplied in the live engine were found by an adversarial audit, conceded and quantified rather than repaired silently (§7.2.4). What the firm acquired is better described as a capability than as a product — sensing demonstrably built and reusable, seizing existing as a committed rule tested once by the content pivot, and reconfiguring designed but not yet performed (§7.4).

The decision. NOT YET. The content pivot has fired and been acted on; the remaining evidence is the efficacy pilot and the legal check; the memo is due in early October, and a documented KILL remains a valid outcome of this dissertation.

10.2 Contribution to practice and to the firm

To practice. The contribution is a worked, publishable example of a corporate-innovation decision in which the belief state is explicit and re-runnable: eight assumptions, each with a prior and a written rationale; every evidence item converted to a likelihood ratio through a published rubric and shrunk by a published quality rule; thresholds fixed before the data; and the arithmetic exposed rather than summarised. The most instructive single episode is that qualitative evidence took the central belief only from 0.50 to about 0.20, and it was the pre-registered survey row that carried it across the pivot line — a concrete answer to the perennial question of how much three interviews should count.

To the firm. Three things. First, a finding the firm would otherwise have discovered after building: its principal asset is the wrong content for the moment of highest need, and the right response is to segment the product by arousal state rather than to abandon the asset or ignore the finding. Second, a portfolio rather than a bet — C3 tests real-music calm demand at almost no cost, C2 hedges consumer-payment friction, C1 is the asset-leverage play, and the gates select among them. Third, a reusable apparatus: a pre-registered survey pipeline, a pilot pipeline with 28 tests, a live decision engine, and a shipped product with research mode built in. The next vertical costs the firm the fieldwork, not the machinery.

10.3 Limitations

- **The survey and focus-group figures are pre-registered pipeline output as published on the project site, not fieldwork.** The results file carries a simulated flag and the focus-group

synthesis is described in the framework as a pre-field template. This is stated wherever those figures appear and is the largest single limitation of the evidence base.

- **The efficacy pilot has not run.** Every efficacy statement in this dissertation is borrowed from meta-analyses conducted in different populations, mostly perioperative, with supervised delivery and GRADE-low trial quality.
- **Non-probability samples throughout**, with hypothetical rather than revealed willingness to pay; the only behavioural instrument, the landing test, has no result.
- **Consumer-grade physiology.** Apple Watch HRV fails equivalence against a chest strap; the live loop is therefore heart-rate-driven, which measures autonomic arousal and is blind to its cause.
- **Insider role.** The founder is the researcher; the controls are declared and implemented, but the conflict is structural rather than removable.
- **Subjective likelihood ratios and an independence assumption.** The likelihood ratios are judgements made against a rubric; the concept joint probabilities multiply correlated beliefs. Both are transparent and contestable — which is their merit, not their defence.
- **Company figures are dated.** Catalogue and stream counts are 2022; the founding date and the artist-relationship scope are unreconciled in the company's own file.
- **Single coder on the netnography**, and the underlying netnography summary was not re-read for this document — only the row count and the patterns already carried in the decision model and the interview synthesis are used.
- **Market sizing does not close.** Two inputs required for a serviceable obtainable market do not exist in the evidence base, so no SOM point estimate is published.

10.4 Further research

1. **Run the comparison the literature has not run.** A properly powered trial of engineered neutral sound against personally selected real music, in acute anxiety, with state anxiety as the primary outcome and residual heart rate as a co-primary. The E1 contrast $\beta 1 - \beta 2$ is a first, underpowered look at exactly this question; nothing in the published literature answers it.
2. **Test whether closed-loop adaptation adds anything over a fixed calming stimulus.** The strong biofeedback evidence is for breathing-paced protocols, not for music-driven adaptation; the generalisation gap should be named and then closed with an adaptive-versus-fixed arm.
3. **Study retention as the dependent variable, not as a footnote.** At a 4.7 per cent thirty-day base rate, retention dominates the unit economics of the entire category; whether anchoring a habit to hardware the user already checks daily changes that rate is an empirical question with a clear design.
4. **Validate the residual as a user-facing quantity.** "Calmer than your usual 22:00 by 9 bpm" is simultaneously an outcome, a reward signal and a piece of interface copy. Whether users read it as credible, and whether reading it changes behaviour, is a distinct study from whether the underlying quantity moves.

5. **Extend the decision apparatus to a second vertical.** The strongest test of the methodological contribution is whether the same register, rubric and gates produce a defensible decision somewhere the answer is not already suspected.

References

APA 7th edition. One alphabetical, de-duplicated list for the whole dissertation, merged from the three chapter lists (`references-ch1478.md`, `references-ch2.md`, `references-ch3569.md`). Where two chapter lists held the same work with different fields, the more completely verified record is used and the choice is recorded in `ASSEMBLY_REPORT.md`. The verification trail is in `sources/verified.md`; fields that no verification pass could establish from a primary or registry source are written in square brackets as "[… to verify]" and are never invented. Their full inventory is in `sources/unverified.md` and in `ASSEMBLY_REPORT.md`.

House decisions applied in this merge (each resolves a divergence between two chapter lists): Coghlan (2019) is single-authored for the fifth edition; Kano et al. (1984) is 14(2), 147–156, per the publisher's J-STAGE record and not the 39–48 that circulates; Bernardi, Porta and Sleight (2006) is the *Heart* paper, the DOI 10.1161/CIRCULATIONAHA.108.806174 belonging to a different and later *Circulation* paper that this dissertation does not cite; Eisenmann, Ries and Dillard is dated **2011**, the publisher's date for note 812-095, not the 2012 SSRN posting date; McGrath and MacMillan (1995) carries a bracketed page range because no primary or index record confirms 44–54; Berger et al. (1993) appeared in the *Center for Quality **of** Management Journal*; Baghai, Coley and White (1999) is the Perseus Books edition; Gibson and Birkinshaw (2004) uses the canonical DOI 10.2307/20159573; and the two Academy of Management articles by Bowman and Hurry (1993) and McGrath (1999) use the Academy of Management DOI form throughout.

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research Policy, 14*(1), 3–22. https://doi.org/10.1016/0048-7333(85)90021-6

Above Avalon. (2025). *Apple Watch* [Analyst notes]. https://www.aboveavalon.com/notes/tag/Apple+Watch

Advertising Law of the People's Republic of China (2015 revision), arts. 17–18, 58–59. https://extranet.who.int/fctcapps/sites/default/files/2024-02/341_%E4%B8%AD%E5%8D%8E%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9B%BD%E5%B9%BF%E5%91%8A%E6%B3%95.pdf

Agrawal, S., & Goyal, N. (2013). Thompson sampling for contextual bandits. [Proceedings, volume and pages to verify]

Ansoff, H. I. (1957). Strategies for diversification. *Harvard Business Review, 35*(5), 113–124.

Apple Inc. (n.d.-a). *HealthKit* [Developer documentation]. https://developer.apple.com/documentation/healthkit [Publication or retrieval date to verify]

Apple Inc. (n.d.-b). *Vitals on Apple Watch* [Support documentation]. https://support.apple.com/guide/watch/vitals-apd15aa7ed96/watchos [Publication or retrieval date to verify]

Apple Watch Series 9 and Ultra 2 heart-rate-variability validation study. (2024). *Sensors*. [Authors, exact title, volume and article number to verify] https://pmc.ncbi.nlm.nih.gov/articles/PMC11478500/

Baghai, M., Coley, S., & White, D. (1999). *The alchemy of growth: Practical insights for building the enduring enterprise*. Perseus Books.

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology,*

34(3), 257–270. https://doi.org/10.1007/s10869-018-9547-8

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Executive, 9*(4), 49–61. https://doi.org/10.5465/ame.1995.9512032192

Behavioral Health Business. (2024, November 20). *Headspace axes 13% of workforce, transitions therapist network to part-time and contract roles*. https://bhbusiness.com/2024/11/20/headspace-axes-13-of-workforce-transition-therapist-network-to-part-time-and-contract-roles/

Berger, C., Blauth, R., Boger, D., Bolster, C., Burchill, G., DuMouchel, W., Pouliot, F., Richter, R., Rubinoff, A., Shen, D., Timko, M., & Walden, D. (1993). Kano's methods for understanding customer-defined quality. *Center for Quality of Management Journal, 2*(4), 3–36. [End page confirmed only from secondary sources; the primary whole-issue scan gives the start page and the twelve-author byline]

Bernardi, L., Porta, C., & Sleight, P. (2006). Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: The importance of silence. *Heart, 92*(4), 445–452. https://doi.org/10.1136/hrt.2005.064600

Blakeley, R. (2024, January 18). Welcome to the sound wellness revolution: Endel's AI-generated soundscapes and the commodification of passive listening. *Musicology Now*. https://musicologynow.org/welcome-to-the-sound-wellness-revolution-endels-ai-generated-soundscapes-and-the-commodification-of-passive-listening/

Bland, D. J., & Osterwalder, A. (2019). *Testing business ideas: A field guide for rapid experimentation*. Wiley.

Blank, S. (2013). Why the lean start-up changes everything. *Harvard Business Review, 91*(5), 63–72.

Bowman, E. H., & Hurry, D. (1993). Strategy through the option lens: An integrated view of resource investments and the incremental-choice process. *Academy of Management Review, 18*(4), 760–782. https://doi.org/10.5465/amr.1993.9402210157

Bradt, J., Dileo, C., & Shim, M. (2013). Music interventions for preoperative anxiety. *Cochrane Database of Systematic Reviews, 2013*(6), Article CD006908. https://doi.org/10.1002/14651858.CD006908.pub2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, T. (2008). Design thinking. *Harvard Business Review, 86*(6), 84–92.

Bundesinstitut für Arzneimittel und Medizinprodukte. (2026). *DiGA-Verzeichnis* [Directory of digital health applications]. https://www.diga-verzeichnis.de/en

Burgelman, R. A. (1983a). A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly, 28*(2), 223–244. https://doi.org/10.2307/2392619

Burgelman, R. A. (1983b). Corporate entrepreneurship and strategic management: Insights from a process study. *Management Science, 29*(12), 1349–1364. https://doi.org/10.1287/mnsc.29.12.1349

Burgelman, R. A. (1984). Designs for corporate entrepreneurship in established firms. *California Management Review, 26*(3), 154–166. https://doi.org/10.2307/41165086

Business of Apps. (2026a). *Calm revenue and usage statistics*. https://www.businessofapps.com/data/calm-statistics/

Business of Apps. (2026b). *Headspace revenue and usage statistics*. https://www.businessofapps.com/data/headspace-statistics/

Calm Health. (2026). *Calm Health* [Company website]. https://health.calm.com/

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science, 66*(2), 564–586. https://doi.org/10.1287/mnsc.2018.3249

Carlgren, L., Rauth, I., & Elmquist, M. (2016). Framing design thinking: The concept in idea and enactment. *Creativity and Innovation Management, 25*(1), 38–57. https://doi.org/10.1111/caim.12153

Chesbrough, H. (2010). Business model innovation: Opportunities and barriers. *Long Range Planning, 43*(2–3), 354–363. https://doi.org/10.1016/j.lrp.2009.07.010

Christensen, C. M. (1997). *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business School Press.

Christensen, C. M., & Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal, 17*(3), 197–218. https://doi.org/10.1002/(SICI)1097-0266(199603)17:3<197::AID-SMJ804>3.0.CO;2-U

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). Know your customers' "jobs to be done." *Harvard Business Review, 94*(9), 54–62.

Christensen, C. M., Raynor, M. E., & McDonald, R. (2015). What is disruptive innovation? *Harvard Business Review, 93*(12), 44–53.

Coghlan, D. (2019). *Doing action research in your own organization* (5th ed.). Sage. [The project's own documents cite this work as "Coghlan & Brannick (2019)"; the fifth edition is single-authored, Brannick having co-authored editions one to four]

Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Sage. [Page range confirmed only from secondary reference lists]

Cooper, R. G. (1990). Stage-gate systems: A new tool for managing new products. *Business Horizons, 33*(3), 44–54. https://doi.org/10.1016/0007-6813(90)90040-I

Cooper, R. G. (2008). Perspective: The Stage-Gate idea-to-launch process—Update, what's new, and NexGen systems. *Journal of Product Innovation Management, 25*(3), 213–232. https://doi.org/10.1111/j.1540-5885.2008.00296.x

Cooper, R. G., & Sommer, A. F. (2016). The Agile–Stage-Gate hybrid model: A promising new approach and a new research opportunity. *Journal of Product Innovation Management, 33*(5), 513–526. https://doi.org/10.1111/jpim.12314

Courtney, H., Kirkland, J., & Viguerie, P. (1997). Strategy under uncertainty. *Harvard Business Review, 75*(6), 66–79. [Authors' full given names are not in the index record and are not asserted; many secondary lists give the first page as 67]

Covin, J. G., & Slevin, D. P. (1991). A conceptual model of entrepreneurship as firm behavior. *Entrepreneurship Theory and Practice, 16*(1), 7–26. https://doi.org/10.1177/104225879101600102

Creswell, J. D., & Goldberg, S. B. (2025). The meditation app revolution. *American Psychologist*. Advance online publication. https://doi.org/10.1037/amp0001576

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Danneels, E. (2002). The dynamics of product innovation and firm competences. *Strategic Management Journal, 23*(12), 1095–1121. https://doi.org/10.1002/smj.275

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Design Council. (n.d.). *Framework for innovation*. Retrieved September 6, 2026, from https://www.designcouncil.org.uk/our-resources/framework-for-innovation/

Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31*(2), 73–86. https://doi.org/10.1016/S0005-7916(00)00012-4

de Witte, M., Spruit, A., van Hooren, S., Moonen, X., & Stams, G.-J. (2020). Effects of music interventions on stress-related outcomes: A systematic review and two meta-analyses. *Health Psychology Review, 14*(2), 294–324. https://doi.org/10.1080/17437199.2019.1627897

Dudík, M., Langford, J., & Li, L. (2011). Doubly robust policy evaluation and learning. [Proceedings, volume and pages to verify]

Easterbrook, J. A. (1959). The effect of emotion on cue utilisation and the organisation of behaviour. [Journal, volume, issue and pages to verify]

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21*(10–11), 1105–1121. https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E

Eisenmann, T. R., Ries, E., & Dillard, S. (2011). *Hypothesis-driven entrepreneurship: The lean startup* (Harvard Business School Background Note No. 812-095). Harvard Business School. [A July 2013 revision is reported in an index snippet only and is not stated here]

Fairfield, T., & Charman, A. E. (2017). Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis, 25*(3), 363–380. https://doi.org/10.1017/pan.2017.14

Federal Trade Commission. (2024, April). *Updated FTC Health Breach Notification Rule puts new provisions in place to protect users of health apps* [Business guidance blog]. https://www.ftc.gov/business-guidance/blog/2024/04/updated-ftc-health-breach-notification-rule-puts-new-provisions-place-protect-users-health-apps

Fierce Biotech. (2023, April). *Prescription app developer Pear Therapeutics files for bankruptcy, lays off staff*. https://www.fiercebiotech.com/medtech/cut-core-prescription-app-developer-pear-therapeutics-files-bankruptcy-lays-staff

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327–358. https://doi.org/10.1037/h0061470

Fortune Business Insights. (2025). *Mental health apps market size, share & global report*. https://www.fortunebusinessinsights.com/mental-health-apps-market-109012

Garcia, R., & Calantone, R. (2002). A critical look at technological innovation typology and innovativeness terminology: A literature review. *Journal of Product Innovation Management, 19*(2), 110–132. https://doi.org/10.1111/1540-5885.1920110

Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. *Academy of Management Journal, 47*(2), 209–226. https://doi.org/10.2307/20159573

Goessl, V. C., Curtiss, J. E., & Hofmann, S. G. (2017). The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis. *Psychological Medicine, 47*(15), 2578–2586. https://doi.org/10.1017/S0033291717001003

Grand View Research. (2025). *Meditation management apps market report*. https://www.grandviewresearch.com/industry-analysis/meditation-management-apps-market-report

Guth, W. D., & Ginsberg, A. (1990). Guest editors' introduction: Corporate entrepreneurship. *Strategic Management Journal, 11*(Special Issue: Corporate Entrepreneurship), 5–15. https://www.jstor.org/stable/2486666 [No DOI exists; the page range comes from the JSTOR landing page and has not been checked against the printed issue]

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ, 336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Harney, C., Johnson, J., Bailes, F., & Havelka, J. (2023). Is music listening an effective intervention for reducing anxiety? A systematic review and meta-analysis of controlled studies. *Musicae Scientiae, 27*(2), 278–298. https://doi.org/10.1177/10298649211046979

Haruvi, A., Kopito, R., Brande-Eilat, N., Kalev, S., Kay, E., & Furman, D. (2022). Measuring and modeling the effect of audio on human focus in everyday environments using brain–computer interface technology. *Frontiers in Computational Neuroscience, 15*, Article 760561. https://doi.org/10.3389/fncom.2021.760561

Hasso Plattner Institute of Design at Stanford. (2018). *Design thinking bootleg*. Stanford d.school. https://dschool.stanford.edu/tools/design-thinking-bootleg

Healthcare Brew. (2025, April 4). *Digital therapeutics Medicare coverage test begins*. https://www.healthcare-brew.com/stories/2025/04/04/digital-therapeutics-medicare-coverage-test-begins

Healthcare Dive. (2021, August). *Headspace, Ginger to merge, creating $3B mental health company*. https://www.healthcaredive.com/news/headspace-ginger-to-merge-create-3b-mental-health-company/605632/

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly, 35*(1), 9–30. https://doi.org/10.2307/2393549

Hernando, D., Roca, S., Sancho, J., Alesanco, Á., & Bailón, R. (2018). [Title to verify]. *Sensors, 18*. [Issue and article number to verify] https://pmc.ncbi.nlm.nih.gov/articles/PMC6111985/

Huang, Y., Wang, Y., Wang, H., Liu, Z., Yu, X., Yan, J., Yu, Y., Kou, C., Xu, X., Lu, J., Wang, Z., He, S., Xu, Y., He, Y., Li, T., Guo, W., Tian, H., Xu, G., Xu, X., … Wu, Y. (2019). Prevalence of mental disorders in China: A cross-sectional epidemiological study. *The Lancet Psychiatry, 6*(3), 211–224. https://doi.org/10.1016/S2215-0366(18)30511-X

Humphreys, M., & Jacobs, A. M. (2015). Mixing methods: A Bayesian approach. *American Political Science Review, 109*(4), 653–673. https://doi.org/10.1017/S0003055415000453

iiMedia Research. (2023). *2023–2024 China sleep economy industry development and consumer demand research report*. https://www.iimedia.cn/c400/92947.html

iiMedia Research. (2025). *2025–2029 China emotion economy consumption trends insight report*. https://www.iimedia.cn/c400/106827.html

Ireland, R. D., Covin, J. G., & Kuratko, D. F. (2009). Conceptualizing corporate entrepreneurship strategy. *Entrepreneurship Theory and Practice, 33*(1), 19–46. https://doi.org/10.1111/j.1540-6520.2008.00279.x

Iskander, N. (2018, September 5). Design thinking is fundamentally conservative and preserves the status quo. *Harvard Business Review*. https://hbr.org/2018/09/design-thinking-is-fundamentally-conservative-and-preserves-the-status-quo

Kahneman, D., Lovallo, D., & Sibony, O. (2011). Before you make that big decision… *Harvard Business Review, 89*(6), 50–60.

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control, 14*(2), 147–156. https://doi.org/10.20684/quality.14.2_147 [The publisher's own J-STAGE archive gives 147–156; the 39–48 range carried in the project's documents and in much secondary literature is a propagated error]

Kanter, R. M. (1985). Supporting innovation and venture development in established companies. *Journal of Business Venturing, 1*(1), 47–60. https://doi.org/10.1016/0883-9026(85)90006-0

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives, 28*(3), 25–48. https://doi.org/10.1257/jep.28.3.25

Kimbell, L. (2011). Rethinking design thinking: Part I. *Design and Culture, 3*(3), 285–306. https://doi.org/10.2752/175470811X13071166525216

Knight, F. H. (1921). *Risk, uncertainty and profit* (Hart, Schaffner & Marx Prize Essays No. 31). Houghton Mifflin. https://archive.org/details/riskuncertaintyp00knigrich

Kozinets, R. V. (2020). *Netnography: The essential guide to qualitative social media research* (3rd ed.). Sage. [SAGE's own pages date the third edition to October 2019 while the repository and most catalogues cite 2020, the copyright year; the repository's year is retained]

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317–325. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Sage. [The publisher's page dates this edition to July 2014; it is universally cited as 2015]

Lassner, A., Siafis, S., Wiese, E., Leucht, S., Metzner, S., Wagner, E., & Hasan, A. (2025). Evidence for music therapy and music medicine in psychiatry: Transdiagnostic meta-review of meta-analyses. *BJPsych Open, 11*(1), Article e4. https://doi.org/10.1192/bjo.2024.826

Latka. (2026). *Calm revenue*. https://getlatka.com/blog/calm-revenue

Liedtka, J. (2015). Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. *Journal of Product Innovation Management, 32*(6), 925–938. https://doi.org/10.1111/jpim.12163

Liedtka, J. (2018). Why design thinking works. *Harvard Business Review, 96*(5), 72–79.

Lovallo, D., & Kahneman, D. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review, 81*(7), 56–63.

Lu, [given name to verify], [remaining authors to verify]. (2021). [Title to verify]. *PeerJ Computer Science, 7*, Article e716. [The repository records the byline only as "Lu, Qiao et al."; whether "Lu Qiao" is one

author or two must be reconciled from the article]

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology, 31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

McCreedy, E. M., [remaining authors to verify]. (2021). [Title to verify]. *Journal of the American Medical Directors Association*. [Volume, issue and pages to verify] https://www.jamda.com/article/S1525-8610(21)01104-X/abstract

McGrath, R. G. (1999). Falling forward: Real options reasoning and entrepreneurial failure. *Academy of Management Review, 24*(1), 13–30. https://doi.org/10.5465/amr.1999.1580438

McGrath, R. G., & MacMillan, I. C. (1995). Discovery-driven planning. *Harvard Business Review, 73*(4), [page range to verify]. https://hbr.org/1995/07/discovery-driven-planning

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153–157. https://doi.org/10.1007/BF02295996

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science, 25*(5), 1106–1115. https://doi.org/10.1177/0956797614524255

Micheli, P., Wilner, S. J. S., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product Innovation Management, 36*(2), 124–148. https://doi.org/10.1111/jpim.12466

Mitrovic, P., & Paladin, A. (2026). Auditory stimuli and heart rate variability: The role of music in cardiovascular regulation. *Frontiers in Cardiovascular Medicine, 13*, Article 1841349. https://doi.org/10.3389/fcvm.2026.1841349

Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Sage. [Edition, series volume and place confirmed only from secondary listings; no primary catalogue record could be retrieved]

Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales? *International Journal of Forecasting, 23*(3), 347–364. https://doi.org/10.1016/j.ijforecast.2007.05.015

Music Business Worldwide. (2023, May). *Universal inks deal with generative AI startup Endel to create "AI-powered, artist-driven functional music"*. https://www.musicbusinessworldwide.com/universal-inks-deal-with-generative-ai-startup-endel-to-create-ai-powered-artist-driven-functional-music/

Nagji, B., & Tuff, G. (2012). Managing your innovation portfolio. *Harvard Business Review, 90*(5), 66–74.

Nakata, C., & Hwang, J. (2020). Design thinking for innovation: Composition, consequence, and contingency. *Journal of Business Research, 118*, 117–128. https://doi.org/10.1016/j.jbusres.2020.06.038

National Institute for Health and Care Excellence. (2023). *Digitally enabled therapies for adults with anxiety disorders: Early value assessment* (HTE9). https://www.nice.org.uk/guidance/hte9

National Institute of Mental Health. (n.d.). *Any anxiety disorder*. https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder [Publication or retrieval date to verify]

Nigg, J. T., [remaining authors to verify]. (2024). [Title to verify]. [Journal, volume, issue and pages to verify] https://pubmed.ncbi.nlm.nih.gov/38428577/

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

O'Daffer, A., Colt, S. F., Wasil, A. R., & Lau, N. (2022). Efficacy and conflicts of interest in randomized controlled trials evaluating Headspace and Calm apps: Systematic review. *JMIR Mental Health, 9*(9), Article e40924. https://doi.org/10.2196/40924

O'Reilly, C. A., III, & Tushman, M. L. (2004). The ambidextrous organization. *Harvard Business Review, 82*(4), 74–81.

O'Reilly, C. A., III, & Tushman, M. L. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior, 28*, 185–206. https://doi.org/10.1016/j.riob.2008.06.002

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Wiley.

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value proposition design: How to create products and services customers want*. Wiley.

Oura Health. (2024). *The Oura API* [Support documentation]. https://support.ouraring.com/hc/en-us/articles/4415266939155-The-Oura-API

Panteleeva, Y., Ceschi, G., Glowinski, D., Courvoisier, D. S., & Grandjean, D. (2018). Music for anxiety? Meta-analysis of anxiety reduction in non-clinical samples. *Psychology of Music, 46*(4), 473–487. https://doi.org/10.1177/0305735617712424

Pedersen, S. K. A., Andersen, P. N., Lugo, R. G., Andreassen, M., & Sütterlin, S. (2017). [Title to verify]. *Frontiers in Psychology, 8*, Article 742. https://pmc.ncbi.nlm.nih.gov/articles/PMC5432607/

Pelletier, C. L. (2004). The effect of music on decreasing arousal due to stress: A meta-analysis. *Journal of Music Therapy, 41*(3), 192–214. https://doi.org/10.1093/jmt/41.3.192

Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review, 84*(1), 62–74.

Pinchot, G., III. (1985). *Intrapreneuring: Why you don't have to leave the corporation to become an entrepreneur*. Harper & Row.

PR Newswire. (2020, June). *MedRhythms receives FDA Breakthrough Device Designation for chronic stroke digital therapeutic* [Press release]. https://www.prnewswire.com/news-releases/medrhythms-receives-fda-breakthrough-device-designation-for-chronic-stroke-digital-therapeutic-301076597.html

PR Newswire. (2023, May 23). *Endel and Universal Music Group to create AI-powered, artist-driven functional music designed to support listener wellness* [Press release]. https://www.prnewswire.com/news-releases/endel-and-universal-music-group-to-create-ai-powered-artist-driven-functional-music-designed-to-support-listener-wellness-301832191.html

PR Newswire. (2025, January 16). *MedRhythms' InTandem rehabilitation system for chronic stroke gait impairment receives final Medicare payment determination from CMS* [Press release]. https://www.prnewswire.com/news-releases/medrhythms-intandem-rehabilitation-system-for-chronic-stroke-gait-impairment-receives-final-medicare-payment-determination-from-cms-302353537.html

Prova Health. (2025). *DiGA reimbursement in Germany: A guide*. https://www.provahealth.com/insights/diga-reimbursement-germany-guide [Exact publication date to verify — the repository records 2024–2025]

Raisch, S., & Birkinshaw, J. (2008). Organizational ambidexterity: Antecedents, outcomes, and moderators. *Journal of Management, 34*(3), 375–409. https://doi.org/10.1177/0149206308316058

Raisch, S., Birkinshaw, J., Probst, G., & Tushman, M. L. (2009). Organizational ambidexterity: Balancing exploitation and exploration for sustained performance. *Organization Science, 20*(4), 685–695. https://doi.org/10.1287/orsc.1090.0428

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Business.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press. [Place of publication not confirmed from a primary source]

Rousseau, D. M. (2006). Is there such a thing as "evidence-based management"? *Academy of Management Review, 31*(2), 256–269. https://doi.org/10.5465/amr.2006.20208679

Sarasvathy, S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review, 26*(2), 243–263. https://doi.org/10.5465/amr.2001.4378020

Sharma, P., & Chrisman, J. J. (1999). Toward a reconciliation of the definitional issues in the field of corporate entrepreneurship. *Entrepreneurship Theory and Practice, 23*(3), 11–28. https://doi.org/10.1177/104225879902300302

Spetzler, C., Winter, H., & Meyer, J. (2016). *Decision quality: Value creation from better business decisions*. John Wiley & Sons.

Spielberger, C. D. (1983). *Manual for the State–Trait Anxiety Inventory (Form Y)*. Consulting Psychologists Press. [The manual is very widely cited with five authors — Spielberger, Gorsuch, Lushene, Vagg and Jacobs — and the byline should be settled against a physical copy; place of publication not confirmed]

Sull, D. N. (2004). Disciplined entrepreneurship. *MIT Sloan Management Review, 46*(1), 71–77.

TDMusic Inc. (2022–2026). *Company background and asset inventory* [Internal documents: business-model deck (2024), thesis tutor presentation (2026), company brochure (2022)]. [Individual document titles, authors and dates to verify; the founding date and the 2022 catalogue figures are flagged for reconciliation in the company's own file]

Tebra. (2025). *Music for mental health* [Healthcare report]. The Intake. https://www.tebra.com/theintake/healthcare-reports/music-for-mental-health

TechCrunch. (2022, April 5). *Endel raises $15M to further develop its AI-powered sound wellness technology*. https://techcrunch.com/2022/04/05/endel-raises-15m-to-further-develop-its-ai-powered-sound-wellness-technology/

TechCrunch. (2025, September 16). *Calm launches standalone iOS app for sleep support*. https://techcrunch.com/2025/09/16/calm-launches-standalone-ios-app-for-sleep-support/

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640

Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43*(2–3), 172–194. https://doi.org/10.1016/j.lrp.2009.07.003

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-

0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. Crown.

36Kr. (2026). 从"香饽饽"到"鸡肋"，睡眠 APP 困于商业化 [From hot commodity to dispensable: Sleep apps stuck on monetisation]. https://36kr.com/p/1852579050655107

TMTPost. (2023, April). [Title to verify — NMPA classification determination for cognitive-function-disorder software]. https://www.tmtpost.com/6473884.html

Toth, A. A., Banks, G. C., Mellor, D., O'Boyle, E. H., Dickson, A., Davis, D. J., DeHaven, A., Bochantin, J., & Borns, J. (2021). Study preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology, 36*(4), 553–571. https://doi.org/10.1007/s10869-020-09695-3

Tracxn. (2026). *Calm* [Company profile]. https://tracxn.com/d/companies/calm/__11__6TilJ32thRjVlCh0DLinrJLZ9o47hTQaLwndAnI

Tushman, M. L., & O'Reilly, C. A., III. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review, 38*(4), 8–29. https://doi.org/10.2307/41165852

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Ulwick, A. W. (2002). Turn customer input into innovation. *Harvard Business Review, 80*(1), 91–97.

U.S. Food and Drug Administration. (n.d.). *General wellness: Policy for low risk devices* [Guidance document]. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-wellness-policy-low-risk-devices [Publication and update dates to verify; the repository records the guidance as cited 2019–2026 with a January 2026 update reported only by secondary law-firm sources]

van Westendorp, P. H. (1976). NSS-Price Sensitivity Meter (PSM): A new approach to study consumer perception of price. In *Proceedings of the 29th ESOMAR Congress* (pp. 139–167). ESOMAR. [Congress ordinal and page range confirmed only from secondary citations; ESOMAR's own archive listing gives neither]

VBData. (2026). [Title to verify — Tide (潮汐) Pre-A funding announcement]. https://www.vbdata.cn/40749

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science, 46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Vickers, A. J., & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. *BMJ, 323*(7321), 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123

Virtual Therapeutics. (2024, July). *Virtual Therapeutics announces results of tender offer to acquire Akili Interactive* [Press release]. Nasdaq. https://www.nasdaq.com/press-release/virtual-therapeutics-announces-results-tender-offer-acquire-akili-interactive-2024-07

Weiser, M., & Brown, J. S. (1996). The coming age of calm technology. [Publication details to verify]

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal, 5*(2), 171–180. https://doi.org/10.1002/smj.4250050207

Woods, K. J. P., Sampaio, G., James, T., Przysinda, E., Cordovez, B., Hewett, A., Spencer, A. E., Morillon, B., & Loui, P. (2024). Rapid modulation in music supports attention in listeners with attentional difficulties. *Communications Biology, 7*(1), Article 1376. https://doi.org/10.1038/s42003-024-07026-3

World Health Organization. (2023). *Anxiety disorders* [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders [The live page now carries an update stamp of 8 September 2025; the figures cited are unchanged, and a dated copy of the version consulted should be retained]

Zott, C., Amit, R., & Massa, L. (2011). The business model: Recent developments and future research. *Journal of Management, 37*(4), 1019–1042. https://doi.org/10.1177/0149206311406265

Appendices

The appendices are artefacts of the project rather than prose, and are not bound into this draft. Each entry below names the artefact, says what it is, and gives the repository path at which it lives, so that a reader or examiner can go to the primary object. Nothing in this index is a finding; every substantive claim drawn from these artefacts is reported, with its provenance, in the chapter that uses it.

Appendix	Artefact	Where it lives
A	Instrument map, and the define-phase artefacts — personas, jobs-to-be-done statements, point-of-view and how-might-we statements, and the territory-scoring matrix. Reported in chapters 4 and 5 as *artefacts*, not as coded findings.	`DESIGN_THINKING_FRAMEWORK.md` §1–5; `research/METHODOLOGY_DESIGN.md` §2
B	Survey v2: the seven-block instrument under the 5W1H frame, with the construct, belief, statistic and likelihood-ratio threshold declared per block, and the pre-registered analysis pipeline.	`surveys/questionnaire-v2.md`; `surveys/analysis/survey_pipeline.py` → `results.json`; published at `site/pages/survey-results.html`
C	Focus-group kit: moderator script, the blind stimulus specifications for clips S-A, S-B and S-C, and the theme synthesis.	`research/FOCUS_GROUP_KIT.md` §1, §3; `research/focus-groups/SYNTHESIS.md`
D	The belief register: eight beliefs with priors and written rationales, the likelihood-ratio calibration rubric, the evidence-quality shrinkage rule, the gate expressions, and the change log.	`research/DECISION_MODEL.md` §1–6; live engine at `site/assets/decision.js`
E	Product blueprint: the design-principles table, screen-by-screen specification, safety rules and the adaptation rule table.	`site/pages/product-design.html` §1, §6.1, §7; `app/ARCHITECTURE.md` §5–9; `app/RECOMMEND_BRIEF.md`
F	Pilot pipeline: the E0 calibration and E1 crossover protocols, the export and analysis code, and the twenty-eight tests that cover the schema, exclusion rule, order inference, injected-effect direction and gate arithmetic.	`research/pilot/README.md` §1–2; `app/tools/export_to_csv.py`; `research/pilot/e1_analysis.py`
G	App screenshots and the release record: web build v0.5.1, TestFlight builds, the App Store 1.0 submission and the Guideline 2.1 reply.	`TODO_2026-08-19.md` §6–16; `app/PLAN.md`
H	Code repository map: the project file map that also generates the public briefing site.	Project file map; briefing site index
I	Verification trail for the bibliography: every work with the record against which its fields were confirmed, and every field that could not be confirmed.	`dissertation/sources/verified.md`; `dissertation/sources/unverified.md`

*Appendix contents are **[to confirm with the author]** for binding: which of A–I are reproduced in full in the submitted document, and which are referenced by repository path only.*