2026 S.T. Yau High School Science Award (Asia)

Research Report

The Team

Registration Number: Phy-233

Name of team member: Edward Jiang
School: Chinese International School
Country: Hong Kong, China

Name of team member: Wai Yan Leung
School: Chinese International School
Country: Hong Kong, China

Name of supervising teacher: Prof. Changyuan Yu
Job Title: DoJJRI & Chair Professor of Photonic Information System
School: The Hong Kong Polytechnic University
Country: Hong Kong, China

Title of Research Report

What Can Three Resonances Reveal?
Experimentally Anchored Inverse Identification of Through-Wall Defects in Aluminum Tubes

Date

17 August 2026

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

What Can Three Resonances Reveal?

Experimentally Anchored Inverse Identification of Through-Wall Defects in Aluminum Tubes

Edward Jiang and Wai Yan Leung

Abstract

A three-resonance measurement carries an information budget rather than one universal accuracy: magnitude, cross-mode pattern, polarization branch, and boundary-condition contrast answer different inverse questions. We mapped that budget using 326 unique real WAV files from impact tests on 36 physical specimens, all 350 mm-long 6063-T5 aluminum tubes. Measured branches establish what the present free–free protocol identifies; supplementary pilots test protocol transfer; calibrated simulations and hybrid inferences ask which added observations could recover information that the measured operator does not contain.

From real paired recordings, 19/21 legacy defects and 18/18 campaign defects exceeded the declared drift-corrected $D = 3$ screening rule, while 5/6 legacy controls were below it in-sample. Leave-one-control-out refits placed 1/6 held-out controls above the rule and produced a 17/21–21/21 range for legacy defects. These counts define a protocol-bound laboratory screening result, with its normalization sensitivity exposed; no contemporaneous campaign controls were recorded. Three damaged circular-hole calibration tubes showed repeatable post-defect orientation-associated branch separation of 7.46 and 21.65 Hz on average in modes 1 and 3. The missing matched intact two-orientation baseline limits causal attribution, but the repeated separation establishes excitation orientation as part of the measurand. A calibrated Timoshenko finite-element model localized twelve real position specimens in the folded free–free coordinate $u = \min(x, L-x)$ with 3.33–3.54 mm mean absolute error across measured-pilot fixed-offset scenarios. Under the declared symmetric three-frequency model, the physical side remains an exactly two-valued inference. An independently calibrated physics–Gaussian-process surrogate predicted six held-out real nominal drill-diameter labels with 0.355–0.432 mm mean absolute error across the same scenarios ($R^2 = 0.946$ without an offset).

The unresolved variables then became experimental-design questions. Real free–free shifts fused with pilot-constrained simulated left- and right-clamp responses selected the correct side in 93.8% of nominal Monte Carlo draws; adding measured between-specimen calibration-residual variance reduced this conditional prediction to 85.5%, with 14.5% mirror flips. This identifies both the potential and the failure mode of an asymmetric-boundary intervention; it is not measured CL/CR fleet reliability. The non-circular after-state experiment did not provide stable four-family separation, so no experimental shape accuracy is assigned after either four-label or three-class grouping. On the same 300 paired simulated 12-specimen LOO draws, a coarse circle / oval / slot taxonomy reached 89.2% modal-only and 96.4% area-informed balanced accuracy, compared with 82.3% and 90.9% for four fabrication labels; normalized anisotropy alone reached 44.1%. Circle / oval / slot is therefore a severity-assisted target for a future validation campaign, while four labels remain a nested stress test. The resulting information map is the central outcome: three resonances support protocol-bound change screening, millimetre-scale folded localization, and held-out nominal-label sizing, while side and detailed contour require a deliberately enriched observation operator.

Keywords: vibration-based damage detection; structural identifiability; Timoshenko beam; modal frequency splitting; Gaussian process; hybrid validation; uncertainty quantification

2026 Si-Yuan High School Science Award
仅用于2026年度科学奖论文公开

Acknowledgements

We are immensely grateful to our mentor, Prof. Changyuan Yu of The Hong Kong Polytechnic University, for the guidance that made the experimental work and repeated recording sessions possible, and to his PhD student, Cui Mingjie, for training us in the use of the machining equipment and supervising the fabrication of the through-wall defects. We also thank the University for providing laboratory space and for the loan of the microphone and USB audio interface.

Our heartfelt thanks go to our parents. Because of them, both of us developed a deep passion for string instruments in childhood through years of lessons and practice — the foundation on which this acoustics paper rests. Nor could it have been done without our friends, who supported us at the nadir of this journey and reminded us that our lives, like a wave, have their own crests and troughs.

Lastly, we would like to thank the S.-T. Yau committee for this wonderful opportunity, and for cultivating a culture of science among the younger generation that will inspire many future changemakers. Along the way we made many memories and met many challenges, and we have undoubtedly grown as individuals.

A detailed account of AI tool usage, including tools, versions, dates, and session counts, is provided in the 'Use of AI tools' section.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Commitments on Academic Honesty and Integrity

We hereby declare that we

1. are fully committed to the principle of honesty, integrity and fair play throughout the competition.
2. perform the research work ourselves and thus truly understand the content of the work.
3. observe the common standard of academic integrity adopted by most journals and degree theses.
4. have declared all the assistance and contribution we have received from any personnel, agency, institution, etc. for the research work.
5. have truthfully disclosed all AI tool usage in the research report and obtained prior approval from supervising teachers. All supporting documentation has been submitted as required. We hereby affirm that no improper use of AI has occurred at any stage of the research process.
6. undertake to avoid getting in touch with assessment panel members in a way that may lead to direct or indirect conflict of interest.
7. undertake to avoid any interaction with assessment panel members that would undermine the neutrality of the panel member and fairness of the assessment process.
8. observe the safety regulations of the venue(s) where we conduct the experiment(s), if applicable.
9. observe all rules and regulations of the competition.
10. agree that the decision of YHSA(Asia) is final in all matters related to the competition.

We understand and agree that failure to honor the above commitments may lead to disqualification from the competition and/or removal of award, if applicable; that any unethical deeds, if found, will be disclosed to the school principal of the team member(s) and relevant parties if deemed necessary; and that the decision of YHSA(Asia) is final and no appeal will be accepted.

Signatures of full team below

Name of team member: Edward Jiang

Name of team member: Wai Yan Leung

Name of supervising teacher: Prof. Changyuan Yu

Noted and endorsed by

Signature of school principal

Date: 26 Aug 2026

Name of school principal: Christine Doleman, Chinese International School

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

Table of Contents

Abstract i

Acknowledgement ii

Commitments on Academic Honesty and Integrity iii

1 Introduction 1

1.1 A simple measurement with a difficult inverse . . . 1
1.2 From frequency shifts to identifiable parameters . . . 1
1.3 Study design and evidence architecture . . . 2
1.4 Questions, contributions, and headline findings . . . 3

2 Theoretical framework 4

2.1 Intact tube as a Timoshenko beam . . . 4
2.2 A dual-effect local defect . . . 5
2.3 First-order modal sensitivity and orientation-dependent branch separation . . . 5
2.4 Detection as distance from a measured control distribution . . . 6
2.5 Position, severity, and the free–free symmetry barrier . . . 6
2.6 Asymmetric boundary conditions as an information intervention . . . 7
2.7 What a shape classifier is allowed to learn . . . 7

3 Methods 7

3.1 Specimens and machined defects . . . 7
3.2 Apparatus and free–free acquisition . . . 7
3.3 Data coverage and analysis scope . . . 9
3.4 Signal processing and peak acceptance . . . 9
3.5 Drift, noise, and detection analysis . . . 10
3.6 Forward-model calibration and real free–free inversion . . . 11
3.7 Independent-group size validation . . . 11
3.8 Pilot-constrained hybrid Multi-BC study . . . 11
3.9 Bayesian posterior check of the information limit . . . 12
3.10 Experimental shape limitation and calibrated sensitivity analysis . . . 12

4 Results 13

4.1 Measured intact dynamics, repeatability, and pilot constraints . . . 13
4.1.1 Intact model and mode acceptance . . . 13
4.1.2 Supplementary microphone and clamp pilots . . . 14
4.2 Measured defect screening under a descriptive rule . . . 14
4.3 Measured post-defect orientation-associated separation in D-C tubes . . . 15
4.4 Measured free–free localization identifies distance, not side . . . 15
4.5 Held-out nominal-diameter label validation . . . 16

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

4.6 Hybrid Multi-BC localization: information gain and fragility . . . 17
4.7 Bayesian posteriors formalize the information limit . . . 18
4.8 Exploratory non-circular trials and controlled sensitivity analysis . . . 20
4.9 Evidence-qualified summary . . . 22

5 Discussion **23**
5.1 The main result is an information budget . . . 23
5.2 Detection: protocol discipline matters more than a perfect point estimate . . . 24
5.3 Post-defect branch separation is both a signal and a warning . . . 24
5.4 Free–free localization is accurate on a quotient space . . . 25
5.5 Sizing: cross-group validation matters more than the choice of regressor . . . 26
5.6 A coarser shape taxonomy clarifies the observation requirement . . . 26
5.7 Model discrepancy is a result, not a nuisance parameter . . . 27
5.8 Validity boundaries and design consequences . . . 27
5.9 Evidence-ordered claims and decisive next experiments . . . 28

6 Conclusion **29**

Author contributions **30**

Data and analysis availability **30**

Use of AI tools **30**

A Evidence scope **30**

B Supplementary materials **31**

References **32**

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 Introduction

1.1 A simple measurement with a difficult inverse

A struck metal tube rings at a small set of reproducible resonance frequencies. Those frequencies can be recorded without a scanning instrument, surface preparation, or access to the defect itself. This makes vibration an attractive first-line screening signal: damage changes mass and stiffness, and therefore changes modal properties wherever the response is measured [1, 2]. Yet the apparent simplicity hides a difficult inverse problem. A natural frequency is a global scalar. It reports that the distribution of mass or stiffness has changed, but does not specify which local change produced it. Measurement drift can be comparable with the effect of a small defect; a symmetric structure can map two different locations to the same spectrum; and a statistical learner can return a precise number even when its training model omits the physics needed for identification.

These distinctions matter because "damage detection" is not one task. Modal-damage reviews distinguish progressively harder identification tasks, from detecting a change to localizing and characterizing it [2]. Each level requires more information than the previous one. A threshold may detect a change without localizing it. Several modes may localize a defect only up to a symmetry. A calibrated forward model may estimate an equivalent severity without distinguishing two geometries that remove the same mechanically effective material. Our study therefore asks not only whether an impact recording can detect a hole, but a sharper question:

> *Which properties of a through-wall defect are identifiable from three bending resonances, and which apparent answers require information that was not measured?*

We address this question with short, thin-walled aluminum tubes because they combine a tractable beam-like global response with a physically important complication: local geometry and pre-existing cross-sectional imperfections can separate otherwise paired bending branches. The resulting orientation-associated separation is not a nuisance to be averaged away. In the damaged D-C tubes it is a repeatable defect-sensitive observation and makes excitation direction part of the measurement definition, although the absence of matched intact two-orientation baselines limits causal attribution of its full magnitude.

1.2 From frequency shifts to identifiable parameters

The classical frequency-shift approach compares measured natural frequencies with an intact baseline and a structural model. Foundational reviews document how changes in modal frequencies, mode shapes, and related model-based quantities have been used for damage detection, location, and characterization [1]. More recent modal reviews emphasize the same practical boundary: small-damage sensitivity, environmental and operational variation, model error, and non-uniqueness remain central problems [2]. Model-based or learned layers do not remove that boundary; their validity still depends on whether the chosen feature is both damage-sensitive and identifiable [3].

Recent theoretical work treats limited spatial information in eigenfrequencies as an experimental-design problem, using stiffness modulation and controlled deformation to resolve symmetry ambiguity in simply supported beams [4]. Our contribution is therefore not the first observation of symmetry-induced non-uniqueness or the first proposal to perturb a structure. It is the experimentally anchored, evidence-tiered map of what a three-resonance aluminum-tube protocol identifies, together with quantified conditional tests of one boundary-condition intervention.

For the present geometry, a Timoshenko beam is a more appropriate low-order model than an Euler–Bernoulli beam because shear deformation and rotary inertia are no longer negligible at the higher

2026 S. T. Yau High School Science Award 仅用于2026年成都中学科学奖公示

measured modes. The shear coefficient is a consequential modeling choice whose calibration can materially affect agreement with beam-vibration data [5]. We use this model not as a claim that a drilled tube is exactly one-dimensional, but as a deliberately compact forward map whose discrepancy can be measured. The defect element removes both stiffness and mass. Those effects oppose each other in the eigenvalue: stiffness loss tends to lower frequency, whereas mass loss tends to raise it. Treating only one effect can therefore reverse or exaggerate a predicted shift.

Axisymmetry adds another layer. Nominally axisymmetric structures possess repeated mode pairs; a local flaw can remove that degeneracy and split their frequencies. Experiments on imperfect cylindrical resonators show that localized mass–stiffness imbalance can split an otherwise paired natural frequency [6]. Our tube modes are global bending modes rather than the resonator mode of that study, so this is a physical analogy, not a substitute forward model. The relevant question is empirical: do repeated tubes exhibit a stable split large enough to exceed experimental drift?

Machine learning enters only after this physics has been specified. Gaussian-process regression (GPR) supplies a flexible probabilistic surrogate for a smooth input–output map [7]. In our workflow, GPR emulates a calibrated physics curve, built from an independent circular-hole group, and is tested on a held-out diameter group. This approach is favored over simply learning from the six validation tubes and predicting them. This separation is essential as a low error would otherwise measure interpolation on reused data rather than external predictive performance.

1.3 Study design and evidence architecture

The study archive contains 326 unique real recordings: 252 from the primary protocol and 74 from earlier pilot and stability sessions. The main corpus contains six controls, twelve fixed-diameter position specimens, six diameter specimens, three orientation-resolved circular-hole specimens, and nine non-circular shape specimens. Measurement coverage is deliberately distinguished by task. Real paired recordings support circular-hole detection, position, diameter, and D-C post-defect orientation-branch analyses. Non-circular after-state trials were conducted, but the tested fixture and acoustic configuration did not yield stable, reproducible four-family separation from the three retained modes. Likewise, no complete specimen-level left- and right-clamp before–after fleet is available for the position study. Seven real Clamp-Left pilot impacts and a six-recording 140-versus-175 mm microphone-position pilot therefore constrain method sensitivity rather than define fleet-level performance.

We therefore separate four evidence classes throughout the paper:

Measured
: quantities computed from real paired main-protocol recordings, including detection, D-C orientation-branch separation, free–free folded localization, and held-out size validation;

Pilot
: supplementary real recordings used to test protocol sensitivity or constrain a model component, but not to estimate fleet-level performance;

Simulated
: outputs of a calibrated physical/statistical model, used for identifiability and sensitivity questions rather than described as experimental accuracy;

Hybrid
: inferences that combine measured free–free shifts with pilot-constrained simulated clamp responses.

This architecture treats uneven measurement coverage as an identification strategy. Measured branches establish which quantities the present observation operator supports; pilot recordings test whether those conclusions survive known protocol changes; calibrated simulations alter one part of the operator at a time;

2026 S. T. Yau High School Science Award 优秀奖 获奖作品学术论文公示

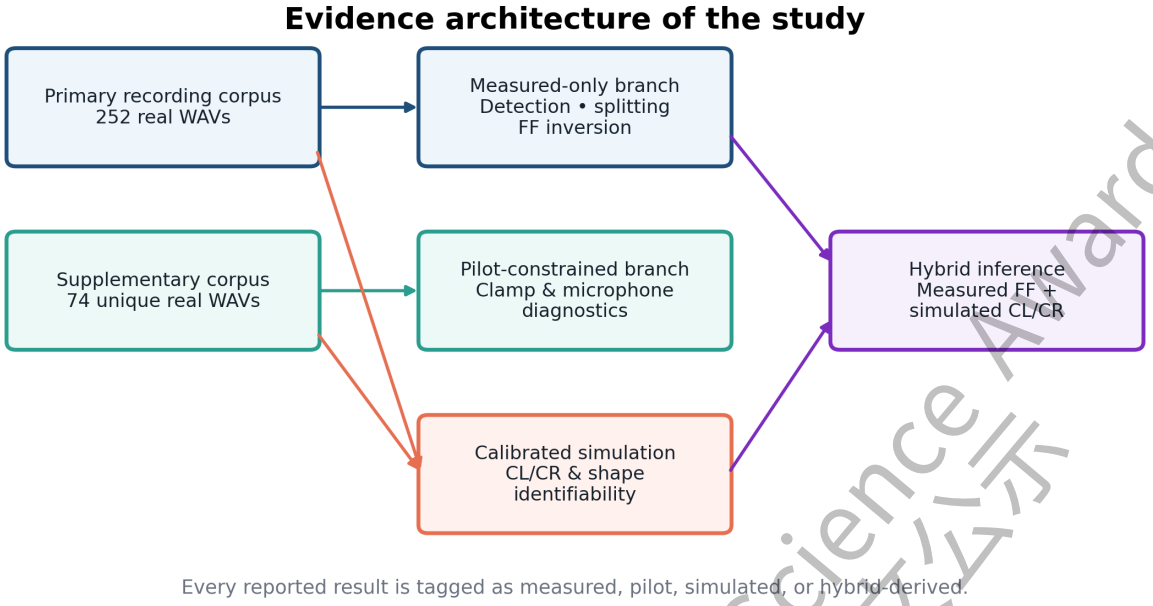


Figure 1. Evidence architecture of the study. The real recordings feed measured and pilot-constrained branches. Simulated CL/CR responses and shape sensitivity analyses are explicitly separated from measured performance. Every quantitative claim in this paper carries one of the four evidence labels defined in the text.

and residual stress tests reveal how the proposed intervention fails. A variable unresolved by the current experiment therefore becomes a specification for the next one. For example, the symmetric free–free model maps mirror positions to the same three-frequency vector, so side recovery requires a symmetry-breaking observation rather than a larger optimizer. Possible side cues in unmodelled waveform or setup asymmetries were not tested.

1.4 Questions, contributions, and headline findings

The study is organized around four fixed questions. First, can three modal shifts detect a through-wall hole relative to a control-derived noise floor? Second, once detection is established, which coordinate of the hole is identifiable under free–free boundary conditions? Third, can an independently calibrated physics–GPR surrogate estimate nominal drill-diameter labels on a held-out real group? Fourth, what additional information would asymmetric boundary conditions or known removed area contribute under realistic uncertainty, and which level of shape taxonomy would that information support?

Our contributions are correspondingly methodological as well as numerical. We (i) document the real recording corpus and its task-specific coverage; (ii) label measurement, pilot constraint, simulation, and hybrid inference at the level of each result; (iii) establish repeatable post-defect orientation-associated bending-branch separation in three independently machined tubes while identifying the intact baseline needed for causal partitioning; (iv) formulate free–free localization on the correct folded coordinate and express mirror non-uniqueness as equal posterior mass; (v) propagate a measured microphone-position pilot and exact leave-one-anchor refits through folded localization and independent-group nominal-label sizing; (vi) quantify an asymmetric-boundary intervention under both nominal and measured-residual-inflated uncertainty; and (vii) use the non-significant shape experiment to compare four fabrication labels with two predeclared three-class taxonomies on identical small-N draws.

The headline findings follow the same information budget. Measured Under the declared drift-corrected screening rule, 19/21 legacy defects and 18/18 campaign defects exceeded $D = 3$, while 5/6 legacy controls were below it in-sample. Leave-one-control-out refits placed 1/6 held-out controls above

2026 年 Tai Yau High School Science Award 应用于 2026 年成相中学科学奖论文公示

the rule and produced 17/21–21/21 legacy defects above it, exposing normalization sensitivity rather than hiding it inside a single rate. Across measured-pilot fixed-offset scenarios, free–free inversion localized the folded coordinate with 3.33–3.54 mm mean absolute error, and independent-group nominal-diameter-label validation gave 0.355–0.432 mm mean absolute error.

HYBRID Pilot-constrained Multi-BC fusion selected the correct mirror side in 93.8% of nominal Monte Carlo draws and 85.5% after measured calibration-residual inflation, which identifies both the intervention's potential and its principal failure mode. MEASURED The non-circular experiment did not provide stable four-family separation, so relabelling does not create an experimental three-class result. SIMULATED On the same 300 paired 12-specimen LOO draws, circle / oval / slot reached 89.2% modal-only and 96.4% area-informed balanced accuracy, compared with 82.3% and 90.9% for four fabrication labels; normalized anisotropy alone remained at 44.1%. The coarse taxonomy is therefore a severity-assisted validation target, while four labels remain a nested stress test.

The rest of the paper develops the forward and observation models, documents measurement coverage, presents results in evidence order, and uses each identifiability boundary to specify the observation needed in a decisive follow-up experiment.

2 Theoretical framework

2.1 Intact tube as a Timoshenko beam

The specimens have length $L = 350\,\mathrm{mm}$, outer diameter $D = 25\,\mathrm{mm}$, and wall thickness $t = 2\,\mathrm{mm}$. With outer and inner radii R_o and R_i, the nominal cross-sectional area and second moment are

$$A_0 = \pi\left(R_o^2 - R_i^2\right), \qquad I_0 = \frac{\pi}{4}\left(R_o^4 - R_i^4\right). \tag{1}$$

For transverse displacement $w(x,t)$ and cross-section rotation $\theta(x,t)$, the Timoshenko strain and kinetic energies may be written

$$U = \frac{1}{2}\int_0^L \left[EI\left(\frac{\partial\theta}{\partial x}\right)^2 + \kappa GA\left(\frac{\partial w}{\partial x} - \theta\right)^2\right] \mathrm{d}x, \tag{2}$$

$$T = \frac{1}{2}\int_0^L \left[\rho A\dot{w}^2 + \rho I\dot{\theta}^2\right] \mathrm{d}x, \tag{3}$$

where $G = E/[2(1+\nu)]$ and $\kappa = 0.604$. The inclusion of rotary inertia and shear deformation is important at modes 2 and 3; the shear coefficient is an explicit modeling choice whose selection can affect agreement with beam-vibration data [5].

The beam is discretized using two-node elements with translational and rotational degrees of freedom. Assembly gives the generalized eigenproblem

$$\boldsymbol{K}\boldsymbol{\phi}_j = \omega_j^2 \boldsymbol{M}\boldsymbol{\phi}_j, \qquad f_j = \frac{\omega_j}{2\pi}. \tag{4}$$

The free–free solver removes the two rigid-body eigenvalues before retaining the first three bending frequencies. The elastic modulus is calibrated once against the measured intact fleet means, with density fixed at $2700\,\mathrm{kg\,m^{-3}}$ and Poisson ratio at 0.33. This gives $E = 67.45$ GPa. Model and measurement agree to within 0.09% at each retained mode, although agreement of intact frequencies alone does not validate the local defect representation.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

2.2 A dual-effect local defect

Drilling changes both sides of the eigenproblem. It reduces local bending stiffness, which lowers frequency, and removes translational and rotary inertia, which raises frequency. We represent a circular through-hole of diameter d centered at x_0 as an axial patch of span d. If element e overlaps this patch by fraction $\eta_e \in [0, 1]$, its properties are modified as

$$(EI)_e = EI_0 \left[1 - \eta_e \gamma_k \left(\frac{d}{D}\right)\right], \tag{5}$$

$$(\rho A)_e = \rho A_0 \left[1 - \eta_e \gamma_m \left(\frac{d}{D}\right)\right], \tag{6}$$

$$(\rho I)_e = \rho I_0 \left[1 - \eta_e \gamma_m \left(\frac{d}{D}\right)\right]. \tag{7}$$

Exact interval overlap is used rather than assigning the entire defect to the nearest element, removing a grid-rounding discontinuity from the forward map. The fitted coefficients are $\gamma_k = 1.265$ and $\gamma_m = 0.242$. They are empirical scale factors for a low-order beam model, not literal fractional losses of the whole tube section. In particular, $\gamma_k > 1$ is possible because a local wall opening changes effective bending transfer more strongly than the axial span ratio alone represents.

The *general forward-model calibration* uses nine real midpoint circular-hole specimens: six C-group tubes spanning 3, 5, and 8 mm, plus three D-C tubes with 5 mm holes. All 27 modal shifts enter a weighted least-squares objective. The resulting $\chi^2 = 110.94$ on 25 degrees of freedom is well above unity per degree of freedom. We retain this discrepancy rather than rescale the noise until the fit appears exact. It records omitted shell deformation, hole-edge geometry, specimen variability, and residual branch-tracking effects. This nine-specimen calibration is used for the general circular-hole and hybrid sensitivity analyses, but *not* for the held-out C-group size test. For that test, a separate two-mode model is re-fitted using only the three D-C specimens ($\gamma_k = 1.447$, $\gamma_m = 0.289$); all six C specimens are excluded from calibration, GPR training, feature scaling, and model selection until final validation.

For the shape identifiability experiment, the circular-hole patch is generalized to a thin-wall arc. Numerical integration over the annulus gives removed area ΔA and the directional second-moment deficit ΔI for a circumferential width, axial length, fill factor, and clocking angle. The simulated element knockdowns then scale with $\Delta I/I_0$ and $\Delta A/A_0$. This model is used only in the SIMULATED branch. It complements the real non-circular trials, which did not yield stable, reproducible four-family separation under the tested fixture and acoustic configuration.

2.3 First-order modal sensitivity and orientation-dependent branch separation

For a small perturbation, the fractional eigenfrequency change can be understood from the Rayleigh quotient:

$$\frac{\delta f_j}{f_j} \approx \frac{1}{2}\left(\frac{\boldsymbol{\phi}_j^{\mathsf{T}} \delta \boldsymbol{K} \boldsymbol{\phi}_j}{\boldsymbol{\phi}_j^{\mathsf{T}} \boldsymbol{K} \boldsymbol{\phi}_j} - \frac{\boldsymbol{\phi}_j^{\mathsf{T}} \delta \boldsymbol{M} \boldsymbol{\phi}_j}{\boldsymbol{\phi}_j^{\mathsf{T}} \boldsymbol{M} \boldsymbol{\phi}_j}\right). \tag{8}$$

The stiffness term is weighted by modal curvature, whereas the mass term is weighted by displacement. Different modes therefore respond differently to the same defect location. This modal fingerprint is the source of position information.

An ideal intact circular tube has equal bending stiffness about any transverse axis, so its two orthogonal bending polarizations are degenerate. A local hole can lift that degeneracy because the stiffness deficit depends on whether the opening lies near the tensile/compressive wall of the bending plane or near its neutral axis; Equation (8) therefore predicts orientation-dependent branch frequencies. Real extruded

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

tubes can also contain pre-existing ovality or seam anisotropy. Experiments on imperfect cylindrical resonators show more generally that local mass–stiffness imbalance can separate paired natural frequencies [6]. The D-C recordings quantify the repeatability and magnitude of the total post-defect separation; without matched intact two-orientation baselines, they do not isolate the drilling contribution uniquely.

2.4 Detection as distance from a measured control distribution

Let $\Delta f_i = f_{i,\mathrm{before}} - f_{i,\mathrm{after}} - \mu_i$ denote the drift-corrected shift in mode i, and let σ_i be the between-session standard deviation derived from paired controls. The detection statistic is

$$D = \left[\sum_{i=1}^{3} \left(\frac{\Delta f_i}{\sigma_i} \right)^2 \right]^{1/2}. \tag{9}$$

The $D > 3$ rule is used as a descriptive control-normalized screening convention, not as an independently calibrated tail-probability test. With only six controls, we do not fit a cross-mode covariance matrix; D is therefore a componentwise scale distance, not a Mahalanobis statistic or calibrated multivariate p-value. We therefore report observed counts together with a leave-one-control-out refit sensitivity: the same controls both define the normalization and provide the in-sample control result. Wilson intervals, where shown, summarize only the observed binomial count and do not include normalization-parameter uncertainty [8].

Equation (9) is insensitive to the signs of the modal shifts. This is useful when uncontrolled impact azimuth changes which polarization branch dominates a legacy recording, but it also means detection can remain strong when the observation is unsuitable for a signed forward-model inversion.

2.5 Position, severity, and the free–free symmetry barrier

For a candidate position x, the calibrated model provides a three-component unit-severity fingerprint $\boldsymbol{s}(x)$. A non-negative specimen-specific severity α is profiled analytically:

$$\hat{\alpha}(x) = \max\left\{0, \frac{\sum_i s_i(x)\Delta f_i/\sigma_i^2}{\sum_i s_i^2(x)/\sigma_i^2}\right\}, \tag{10}$$

with profile objective

$$\chi^2(x) = \sum_i \frac{[\Delta f_i - \hat{\alpha}(x) s_i(x)]^2}{\sigma_i^2}. \tag{11}$$

For a uniform free–free beam, reflection about the midpoint leaves each modal energy density unchanged. Hence

$$\boldsymbol{s}_{\mathrm{FF}}(x) = \boldsymbol{s}_{\mathrm{FF}}(L - x). \tag{12}$$

Under this symmetric free–free frequency-only model, the two positions are not merely difficult to distinguish; they are the same point in the retained three-frequency observation space. The identifiable coordinate is therefore

$$u = \min(x, L - x), \qquad 0 \le u \le L/2. \tag{13}$$

All measured free–free localization errors are evaluated in u. Selecting the candidate nearest the labelled end after seeing the ground truth would be a mirror correction, not an inference. Real suspension, impact, microphone, or manufacturing asymmetries might imprint additional waveform information, but that information was neither represented by the retained frequency features nor validated here.

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

2.6 Asymmetric boundary conditions as an information intervention

Clamping one end changes the modal energy densities and destroys the reflection symmetry in equation (12). The supplemental CL pilot shows that the clamp is far from rigid: its repeatable first mode is approximately 140.2 Hz. We represent the root by a rotational spring. A measured-pilot calibration gives $k_\theta \approx 6.56 \times 10^3\,\mathrm{N\,m\,rad^{-1}}$; higher clamp modes are less stable and receive larger uncertainty in the hybrid analysis.

For a physical defect coordinate x, the left-clamped model uses the coordinate $x - 20\,\mathrm{mm}$ along the 330 mm effective beam, whereas the right-clamped model uses $330\,\mathrm{mm} - x$. The combined fingerprint

$$\boldsymbol{s}_{\mathrm{MB}}(x) = \begin{bmatrix} \boldsymbol{s}_{\mathrm{FF}}(x) \\ \boldsymbol{s}_{\mathrm{CL}}(x) \\ \boldsymbol{s}_{\mathrm{CR}}(x) \end{bmatrix} \tag{14}$$

is generally different at x and $L - x$. The present dataset does not contain the required paired CL/CR fleet recordings, so this symmetry-breaking advantage is evaluated as a HYBRID Monte Carlo result, not an experimental localization score.

2.7 What a shape classifier is allowed to learn

The shape simulation produces two physically distinct types of feature. Raw sensitive-branch severity is strongly influenced by how much material was removed. A normalized anisotropy feature, $q_3 = s_3/A$, divides the third-mode branch split s_3 by removed area A and is intended to reduce this trivial size cue. We evaluate both a modal-only feature set and a feature set augmented with independently known removed area. If the augmented classifier is substantially better, the improvement should be attributed to the extra geometric measurement, not to the modal spectrum. This distinction is central to the interpretation of the shape results.

3 Methods

3.1 Specimens and machined defects

Thirty-six commercial 6063-T5 aluminum tubes were cut to 350 mm. Nominal outer diameter was 25.0 mm and wall thickness 2.0 mm. The laboratory protocol records visual inspection, caliper checks, and a fixed left-end axial/circumferential datum before specimen coding. Through-wall features were machined through one wall at a common circumferential datum and deburred. Axial coordinates are measured from the labelled left end.

The B positions were selected to sample distinct combinations of the first three modal energy densities. The C and D defects lie at the midpoint, separating size/shape questions from position. Specimen labels, recording conditions, and trial metadata were retained throughout the analysis.

3.2 Apparatus and free–free acquisition

The recorded apparatus used a 10 mm steel-ball pendulum on a 300 mm string released from a fixed angle to produce repeatable impacts. An omnidirectional Behringer ECM8000 condenser microphone at approximately 50 mm standoff fed a Behringer UMC202HD interface. Audacity recorded mono WAV files at 96 kHz, 24-bit, with automatic gain control, normalization, compression, and equalization disabled.

For the free–free configuration, the tube was suspended on nylon loops near the first-mode nodes at

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Table 1. Physical specimen matrix. The non-circular after-state trials were conducted, but their three-mode responses did not yield stable, reproducible four-family discrimination under the tested fixture and acoustic configuration. Dimensions for non-circular features are circumferential width × axial length.

Group	N	Geometry and station	Role in this analysis
A	3	intact	control and between-session variation
E	3	sham-handled, no material removed	independent control handling
B	12	5 mm circular holes at 70, 120, 175, or 245 mm; three per station	detection and free–free folded localization
C	6	midpoint circular holes of 3, 5, or 8 mm; two per diameter	held-out real size validation
D-C	3	midpoint 5 mm circular holes	real post-defect 0°/90° branch-separation anchor and independent size-model calibration
D-OT	3	midpoint 8×4 mm transverse ovals	experimental after-state response non-significant; sensitivity analysis
D-OA	3	midpoint 4×8 mm axial ovals	experimental after-state response non-significant; sensitivity analysis
D-SA	3	midpoint 3×15 mm axial slots	experimental after-state response non-significant; sensitivity analysis

approximately 78 and 272 mm. The pendulum contacted an end-region antinode, and three impacts were saved as separate files for each available specimen/state combination. The median of the three extracted frequencies is the specimen-level observation. A complete specimen-level CL/CR before–after fleet was not available for the position analysis, so asymmetric-boundary performance is evaluated as a pilot-constrained sensitivity study rather than a fleet-level experimental score.

Two free–free after-state protocols are present. The *legacy* recordings used a 175 mm microphone position and did not consistently control the impact azimuth relative to the defect. The later *campaign* recordings used a 140 mm microphone position and a controlled orthogonal impact orientation for the B and C groups. A six-WAV real microphone-position pilot measured mode-1/2/3 mean $f_{140} - f_{175}$ offsets of −0.009, +0.265, and +1.577 Hz; the corresponding paired-difference standard deviations were 0.014, 0.232, and 0.598 Hz. The larger mode-3 difference is evidence that extracted peaks are not perfectly invariant to observation geometry.

To quantify rather than merely note this protocol mismatch, four fixed offset vectors were defined before evaluation: zero, the measured pilot mean, and the pilot mean minus or plus one paired-difference standard deviation in each mode. Each vector was applied to the existing campaign before–after shifts, after which the unchanged B folded-coordinate localization and held-out C nominal-label pipelines were rerun. These are deterministic protocol-transfer sensitivity scenarios conditioned on a small pilot. The mean±SD rows are not confidence intervals, additional specimens, or an empirical fleet-wide distribution,

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Table 2. Data coverage for the 326 unique real WAV files. Primary recordings support the reported performance analyses; supplementary data constrain or test methods but do not enlarge the denominator of a headline performance claim.

Evidence role	WAVs	Boundary	Permitted use
Primary protocol	252	FF	detection, measured post-defect branch separation, folded localization, real size validation
Clamp calibration anchor	7	CL	first-mode repeatability and finite-root calibration
Microphone method validation	6	FF	quantify 140-versus-175 mm frequency offsets
Secondary measured pairs	6	FF	qualitative early-pair support only
Session-stability anchors	45	FF	cross-session diagnostic; not primary performance
Exploratory only	10	mixed	appendix/protocol diagnosis only

and no scenario was selected after inspecting which produced the most favorable B/C error.

The D-C circular-hole tubes have real after-state recordings at two azimuths. The 0° and 90° labels refer to orthogonal bending planes relative to the defect clocking. Three impacts per branch were recorded. The branch difference is computed within the same specimen and after-state, reducing sensitivity to between-specimen variation. Matched intact recordings at both orientations were not retained, so this design measures total post-defect orientation-associated separation but cannot partition it uniquely into drilling-induced and pre-existing tube-anisotropy components. Non-circular after-state trials were also conducted, but their three-mode responses were not sufficiently stable and distinguishable under the tested fixture, environmental, and microphone conditions to support a reproducible experimental shape-classification accuracy.

3.3 Data coverage and analysis scope

The primary data set contains 252 unique free–free recordings: 108 intact recordings (36 specimens × three trials), 90 legacy after-state recordings, and 54 orientation-controlled campaign after-state recordings. The legacy set includes three trials for six controls, twelve B tubes, six C tubes, and two orientations for three D-C tubes. The campaign contains three trials for the twelve B and six C specimens.

Supplementary recordings include seven Clamp-Left pilot impacts, six microphone-position tests, six early free–free before–after pairs, earlier-session intact recordings, and exploratory fixture/defect trials. These recordings support protocol checks and pilot constraints, but the primary performance denominators remain specimen based.

For each recording, the study records specimen identity, condition, boundary condition, microphone position, orientation, trial, and the three extracted frequencies. These metadata are used to preserve protocol distinctions throughout the analysis.

3.4 Signal processing and peak acceptance

The first 5 ms of each recording is removed to avoid the contact transient. A Hann window of $N = 131{,}072$ samples (1.365 s) is applied and zero-padded to $N_{\mathrm{FFT}} = 2^{20}$. At 96 kHz, the displayed frequency grid is approximately 0.092 Hz. Within mode-specific windows around 1.15, 3.03, and 5.57 kHz, the local maximum is refined by three-point parabolic interpolation on log magnitude. If α, β, and γ denote adjacent log-magnitudes around peak bin k^*, the sub-bin correction is

$$\delta = \frac{1}{2}\frac{\alpha - \gamma}{\alpha - 2\beta + \gamma}, \qquad \hat{f} = (k^* + \delta)\Delta f_{\mathrm{bin}}. \tag{15}$$

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

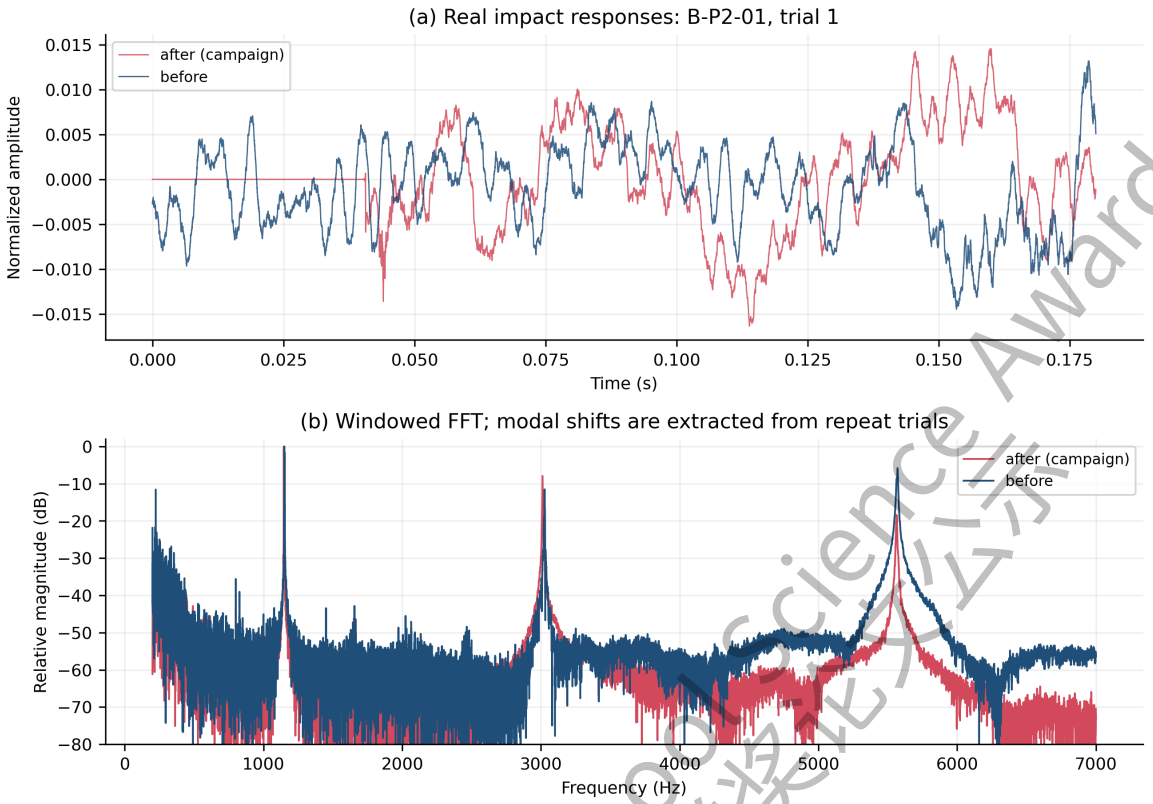


Figure 2. Representative real free–free impact responses for B-P2-01 before and in the orientation-controlled campaign after-state. Time amplitudes are normalized by each full recording's maximum. The FFT is shown in relative decibels; all three retained peaks are resolved. The plot demonstrates the recorded signal, not a synthetic waveform.

The same procedure is used for before and after files. No sign-based rejection is imposed: the dual mass–stiffness model permits small or negative shifts in some mode/position combinations.

Mode 4 was examined and rejected before the final analyses. Its median within-session standard deviation was 5.033 Hz, compared with 0.012, 0.082, and 0.415 Hz for modes 1–3, and controls produced apparent mode-4 between-session shifts up to approximately 78 Hz. The three retained modes therefore define both the observation vector and the three degrees of freedom of the detection statistic.

3.5 Drift, noise, and detection analysis

The paired controls A and E define the free–free noise floor. A-01 produced the anomalous paired-control excursion. The declared drift-center convention therefore uses the mean shift of the other five controls, giving

$$\boldsymbol{\mu} = (-0.155, -0.489, -0.898)\ \text{Hz}. \tag{16}$$

The componentwise standard deviations conservatively retain A-01 and are

$$\boldsymbol{\sigma}_{\text{FF}} = (0.425, 0.878, 1.019)\ \text{Hz}. \tag{17}$$

No cross-mode covariance is fitted from this six-control sample; equation (9) uses only these component-wise scales. The drift-corrected median trial shifts enter equation (9); a specimen is recorded as above the descriptive screening rule when $D > 3$. The 21 legacy circular-hole specimens (B, C, and D-C) and six A/E controls define the observed legacy count. Because the normalization convention is determined from

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

the same six control records—five entering $\boldsymbol{\mu}$, all six entering $\boldsymbol{\sigma}_{\mathrm{FF}}$—and those records are then screened, this is an in-sample, control-normalized analysis rather than an independently calibrated sensitivity or specificity estimate. We therefore also refit the center and spread after holding out each control in turn and report both the held-out-control count above threshold and the range of legacy defects above threshold. Campaign reporting uses 18 defective B/C specimens; no A/E control was recorded in that session, so campaign specificity is undefined. Wilson intervals, where reported, are conditional summaries of observed proportions only and do not include uncertainty from estimating the normalization parameters.

3.6 Forward-model calibration and real free–free inversion

The intact model uses 140 equal elements for the 350 mm free–free tube. E is fitted to the three measured intact fleet means. For the general localization/hybrid model, the circular-defect coefficients γ_k and γ_m are fitted jointly using modes 1–3 of the six C campaign specimens and the sensitive branch of the three D-C specimens. The objective uses $\boldsymbol{\sigma}_{\mathrm{FF}}$ and has 27 scalar observations, two fitted coefficients, and 25 degrees of freedom.

For B-group localization, a 5 mm-diameter sensitivity curve is evaluated at 2.5 mm increments over $x \in [20, 330]$ mm. Equations (10) and (11) profile severity and score each position. The estimator is evaluated only in the folded coordinate u of equation (13). The primary metrics are mean and median $|\hat{u} - u|$ over all twelve real B campaign specimens. No post hoc choice between mirror positions is applied.

3.7 Independent-group size validation

Sizing is intentionally recalibrated to prevent leakage from the validation group. Only the three real D-C 5 mm specimens and their 90° sensitive branches fit a separate two-mode (modes 1 and 3) circular-defect curve, giving $\gamma_k = 1.447$ and $\gamma_m = 0.289$ for the size analysis. These coefficients replace, rather than inherit, the nine-specimen general calibration in this branch. Mode 2 is retained as a diagnostic but excluded from this midpoint size calibration because it carries little stable sensitivity.

The D-C-calibrated model is evaluated on a dense diameter grid from 1.5 to 9.5 mm in 0.1 mm steps. Each physics vector is projected by weighted least squares onto the 5 mm reference, yielding a one-dimensional monotone severity coordinate. A standardized GPR with a constant-times-radial-basis-function kernel plus a small white-noise term maps this coordinate to diameter. The six C campaign specimens (two each at 3, 5, and 8 mm) are excluded from coefficient calibration, feature scaling, GPR training, kernel selection, and model selection, then used once as an external group-level test. We report MAE, median absolute error, and R^2. The GPR's internal interpolation standard deviation is archived but not presented as total experimental uncertainty.

To test calibration leverage without changing either validation group, the general location calibration is refitted after omitting each of its 9 C/D-C anchor specimens in turn, while the original twelve B evaluations remain fixed. The independent sizing calibration is likewise refitted after each one-specimen omission from its 3 D-C anchors, with all six C specimens still held out. For each full fit, a local Laplace approximation is formed from the finite-difference Hessian of the weighted objective, and the resulting coefficient covariance is propagated through the frozen inverse map. These calculations isolate sensitivity to the retained calibration anchors and local coefficient covariance; they do not absorb the observed model residual, protocol transfer, or structural alternatives outside the low-order forward model.

3.8 Pilot-constrained hybrid Multi-BC study

Seven real CL pilot impacts provide the only direct clamp recording set. Their first-mode median is 140.2 Hz with approximately 0.13 Hz standard deviation in the consensus analysis. A finite-root cantilever

2026 S-T Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

model is fitted to this pilot, producing $k_\theta \approx 6.56 \times 10^3\,\mathrm{N\,m\,rad^{-1}}$. Candidate higher modes near 1.09 and 2.72 kHz are less repeatable, so the hybrid noise vector is deliberately widened to

$$\boldsymbol{\sigma}_{\mathrm{CL/CR}} = (0.30, 1.80, 3.00)\ \mathrm{Hz}. \tag{18}$$

For each B specimen, the real campaign FF shift fixes a specimen-specific severity scalar. The model then generates expected CL and CR shifts at the known diameter but does not use the true side during inversion. In each of 3,000 Monte Carlo draws per specimen, CL and CR vectors receive independent Gaussian measurement noise and a 10% multiplicative model perturbation. The nine-component FF/CL/CR profile search is repeated on the same 2.5 mm grid. We record conditional correct-side probability, median total absolute error, error conditional on selecting the correct side, and a profile-score ratio between the best two sides; none of these outputs is presented as a measured reliability rate.

The sensitivity scenarios vary clamp noise and model error from optimistic pilot conditions to conservative pilot-constrained conditions. A separate nominal scenario adds the between-specimen standard deviation of the measured calibration residual to both FF and simulated clamp variance components. These scenarios are conditional design analyses, not estimates of deployment performance. The measured pilot constrains the order of magnitude of one left-root condition; no Clamp-Right pilot and no empirical distribution of clamp-model form error were available. It therefore does not validate the simulated B-fleet CL/CR response specimen by specimen.

3.9 Bayesian posterior check of the information limit

For an uncertainty-aware statement of the same inverse problem, we evaluate a grid posterior over physical position $x \in [20, 330]$ mm and circular-defect severity $\alpha \in [0, 3]$, with uniform priors. At every grid point, the likelihood is Gaussian in the retained frequency-shift vector, using the declared FF control scale for the measured FF posterior. We marginalize the normalized two-dimensional posterior over α to report posterior mass in the left and right halves of the tube. This calculation uses the same forward model and real B-group FF observations as the profile search; it adds a probability distribution, not new experimental information.

A second, explicitly conditional posterior combines the real FF vector with nominal CL and CR response vectors generated by the pilot-constrained finite-root model at the specimen's recorded diameter and physical position. The simulated CL/CR vectors use the nominal variance of the Hybrid study. This posterior is a simulated experiment that asks how much side information the specified asymmetric-boundary intervention could supply. Its plotted representative posterior is a nominal-noise calculation and is interpreted jointly with the residual-inflated Monte Carlo scenario. It is not a posterior from a measured CL/CR specimen fleet and is labelled Hybrid throughout.

3.10 Experimental shape limitation and calibrated sensitivity analysis

Non-circular after-state trials were carried out for the D-OT, D-OA, and D-SA families. In the tested measurement configuration, their three-mode signatures did not provide a stable, repeatable separation across the planned four fabrication families. This empirical limitation is not treated as evidence that the geometries are identical; it shows that the available fixture, acoustic environment, microphone chain, and global-frequency features did not provide sufficient discrimination in that configuration. We therefore report neither an experimental four-family accuracy nor an experimental accuracy after relabelling the same tubes into three classes. Instead, the calibrated analysis below asks a complementary design question: under controlled variation in machining and frequency noise, which taxonomy and which combi-

2026 S. T. Yau High School Science Award 应用于2026丘成桐中学科学奖论文公示

nation of geometric and modal information would be sufficiently observable to justify a future validation experiment?

The sensitivity analysis is anchored to the three real D-C specimens, for which both polarization branches exist. Modes 1 and 3 from both branches fit the shape-model stiffness and mass scales. Nominal geometries are a 5 mm circle, 8×4 mm transverse oval, 4×8 mm axial oval, and 3×15 mm axial slot. Manufacturing variation perturbs circumferential width and axial length; local response derivatives are evaluated by centered 0.1 mm steps. Between-specimen noise is anchored to the observed D-C sensitive-branch and split variability, with a lower bound of one quarter of the corresponding FF control sigma.

For the large-sample four-label sensitivity map, each original fabrication family contributes 200 simulated training specimens and at least 150 independently generated test specimens. A standardized shrinkage linear discriminant analysis (LDA) is repeated 20 times for each combination of machining standard deviation (0.05, 0.10, 0.20, or 0.30 mm) and noise multiplier (0.5, 1, or 2). The nominal condition is 0.10 mm machining SD and unit noise.

The actual planned experiment had only three specimens per fabrication family. To reflect that design, 300 independent 12-specimen datasets are sampled from the nominal 800-specimen simulation pool, each containing three examples of D-C, D-OT, D-OA, and D-SA. Every sampled specimen is retained while the identical draw is evaluated under three predeclared label maps: (i) the four nominal fabrication labels; (ii) three geometry families—circle, pooled oval, and slot; and (iii) three spectrally motivated families—circle, transverse oval, and pooled axial-elongated. The second map asks a coarser contour question, whereas the third tests the near-degeneracy of D-OA and D-SA after area normalization; neither map is chosen by inspecting the resulting score.

Full leave-one-out validation is repeated for three precisely defined feature sets:

$$\begin{aligned} \texttt{modal-only} &: [\Delta f_{1,\mathrm{sens}}, \Delta f_{3,\mathrm{sens}}, s_1, s_3], \\ \texttt{normalized-anisotropy} &: [q_1, q_3], \\ \texttt{area-informed} &: [\|\Delta \boldsymbol{f}_{\mathrm{sens}}\|, s_1, s_3, q_1, q_3]. \end{aligned}$$

Here $q_i = s_i/A$ requires independently known removed area A, but area itself is not supplied directly as a classifier feature. In every LOO fold, the scaler and shrinkage LDA are fitted on the 11 training specimens only, with equal class priors. Because pooling creates unequal class counts while all 12 specimens remain in the test design, we report both overall and balanced accuracy, the distribution of correct classifications out of 12, and aggregated class recalls. The comparison is paired across taxonomies; it does not obtain a larger score by discarding a difficult specimen or selecting a favorable split.

4 Results

4.1 Measured intact dynamics, repeatability, and pilot constraints

4.1.1 Intact model and mode acceptance

The measured intact fleet means for modes 1–3 were 1153.23, 3028.16, and 5573.90 Hz. At the calibrated $E = 67.45$ GPa, the Timoshenko model predicted 1154.24, 3028.17, and 5568.97 Hz, corresponding to relative deviations of 0.088%, 0.001%, and −0.088%. This symmetric agreement validates the intact global stiffness scale, but not the local defect element.

Within-session median standard deviations rose strongly with mode: 0.012, 0.082, 0.415, and 5.033 Hz for modes 1–4. The fourth mode was therefore omitted. This exclusion is consequential: adding a nominally more informative frequency would instead inject peak-identity error much larger than the

2026 S. T. Yau High School Science Award 应用于2026丘成桐中学生数学奖公示

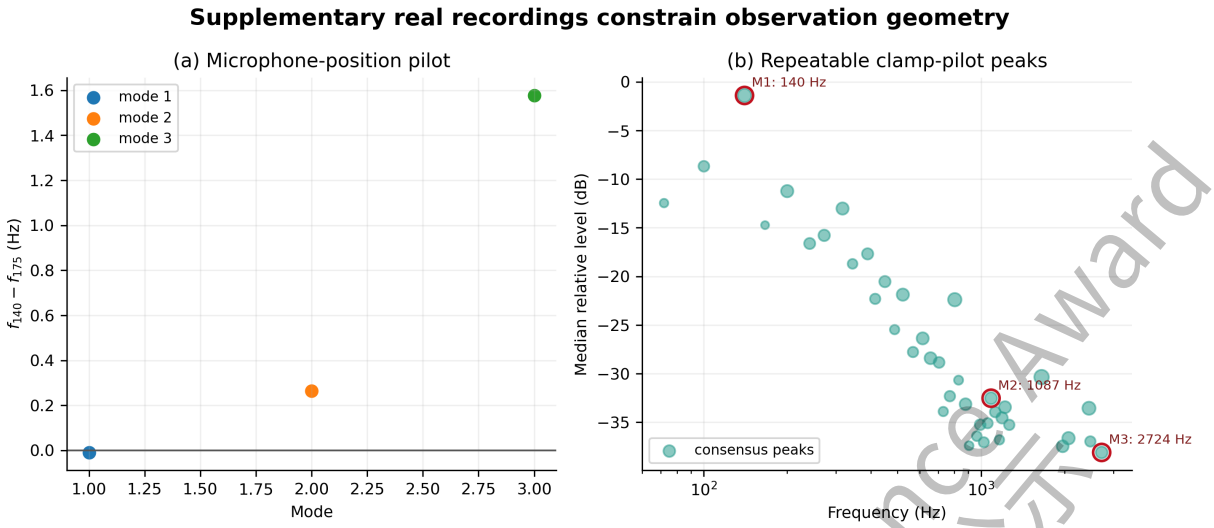


Figure 3. Pilot Supplementary real recordings. (a) Mean extracted-frequency difference between microphone positions 140 and 175 mm. (b) Consensus peaks across seven real left-clamp pilot impacts; point size encodes spectral prominence and red circles mark the three model candidates. The first candidate is highly repeatable, whereas the higher assignments are used only as weak constraints.

defect signal.

The circular-defect calibration gave $\gamma_k = 1.265$ and $\gamma_m = 0.242$. The weighted residual was $\chi^2 = 110.94$ on 25 degrees of freedom. Thus, the model captured the main position/diameter trend but not the observations to the control-derived noise level. We treat the excess residual as model discrepancy in every later interpretation.

4.1.2 Supplementary microphone and clamp pilots

Pilot The 140-versus-175 mm microphone comparison gave mode-1/2/3 mean extracted-peak offsets of -0.009, $+0.265$, and $+1.577$ Hz; the corresponding paired-difference standard deviations were 0.014, 0.232, and 0.598 Hz. These are setup-dependent extraction differences rather than evidence that microphone position changes a structural eigenfrequency. Their propagation through the existing B and C pipelines is reported as fixed protocol-transfer scenarios in table 4.

The seven real CL pilot impacts produced a repeatable candidate at 140.2 Hz; the run-to-run standard deviation of the first-mode assignment was approximately 0.13 Hz. Higher candidates around 1.09 and 2.72 kHz were appreciably less stable. A finite rotational-root fit reproduced the first mode at $k_\theta \approx 6.56 \times 10^3\,\mathrm{N\,m\,rad^{-1}}$, but its predictions for the two higher candidates remained imperfect. This is why the hybrid CL/CR uncertainty grows with mode rather than reusing the FF control sigmas.

4.2 Measured defect screening under a descriptive rule

Measured Under the declared drift-corrected statistic, 19/21 legacy defective circular-hole specimens exceeded $D = 3$, while 5/6 controls were below the same rule in-sample; A-01 remained above it. Because these six controls also define the drift and spread normalization, neither count is an independently calibrated diagnostic sensitivity or specificity estimate. A leave-one-control-out refit makes that dependence visible: 1/6 held-out controls exceeded the rule, and 17/21–21/21 legacy defects exceeded it across the six refits.

All 18/18 B/C campaign defects exceeded $D = 3$. No contemporaneous campaign controls were recorded, and the campaign used a different microphone geometry from the legacy protocol. This is

2026 Siyuan High School Science Award 仅用于2026年思源中学科学奖论文公示

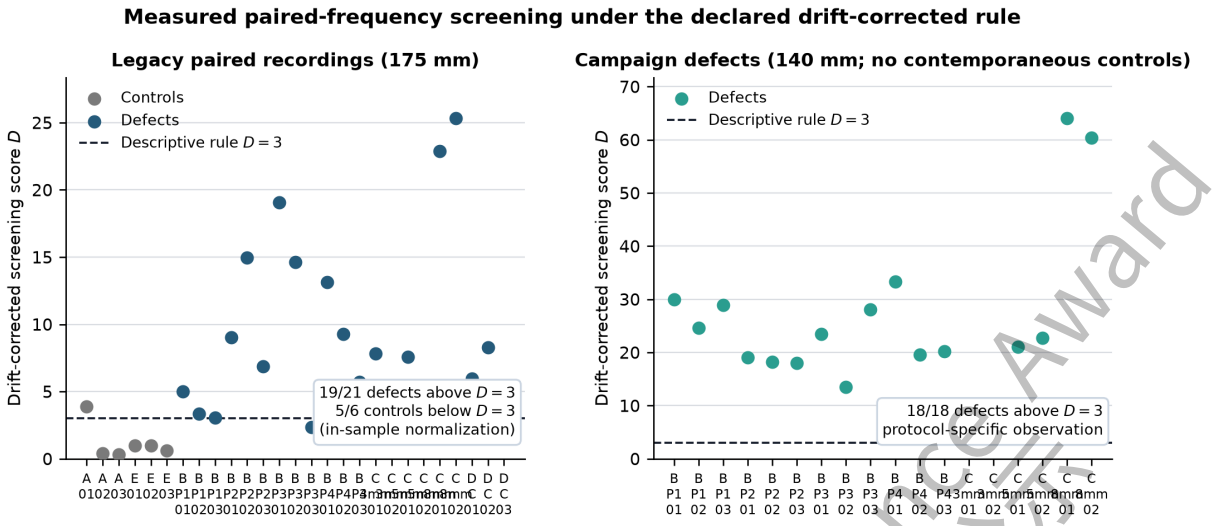


Figure 4. MEASURED Paired-frequency screening under the declared drift-corrected rule. The legacy panel includes all six controls and 21 defective circular-hole specimens; the campaign panel contains only 18 defective B/C specimens. The dashed $D = 3$ line is a descriptive control-normalized screening convention, not an independently calibrated tail-probability threshold.

therefore a protocol-specific observed screening result, not a transferable estimate of sensitivity, specificity, or overall accuracy. The stronger separation from the rule is compatible with improved polarization control, but matched intact transfer standards would be required to isolate that contribution.

The protocol comparison also clarifies why the same statistic can be useful for screening but insufficient for inversion. D uses squared standardized shifts and is unaffected by a sign reversal. A forward fingerprint depends on the signed vector. Uncontrolled polarization can therefore leave D large while moving the observation toward the wrong branch of the inverse map.

4.3 Measured post-defect orientation-associated separation in D-C tubes

MEASURED The three D-C tubes exhibit repeatable post-defect orientation-associated separation between the two bending branches; because matched intact two-orientation baselines were not retained, we do not attribute the full observed separation uniquely to drilling. Defining $s_i = f_{i,0^\circ} - f_{i,90^\circ}$, with the 90° branch used as the sensitive branch in later calibration, mean signed separations were 7.46 Hz in mode 1, −3.40 Hz in mode 2, and 21.65 Hz in mode 3. The specimen ranges were 6.75–8.65, −4.86–−1.73, and 20.19–24.27 Hz, respectively. The 90° sensitive branch produced mean paired shifts of 7.28, 0.88, and 19.14 Hz.

The sign and magnitude of mode 2 differ from modes 1 and 3. We do not force all three modes into one qualitative story: the perturbation depends on the competing stiffness and mass terms as well as the local modal fields. What is robust is the repeated doublet separation itself and the much stronger orientation dependence in modes 1 and 3.

4.4 Measured free–free localization identifies distance, not side

MEASURED The free–free profile search was evaluated on all twelve B campaign specimens. With no microphone-position offset applied, the folded-coordinate mean absolute error was 3.33 mm, the median was 3.75 mm, and the maximum was 7.5 mm. Across the zero, pilot-mean, and pilot-mean±SD fixed-offset scenarios, folded-coordinate MAE remained within 3.33–3.54 mm, while maximum absolute error ranged from 7.5–10.0 mm. The largest numerical difference between mirror minima in the no-offset computation

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

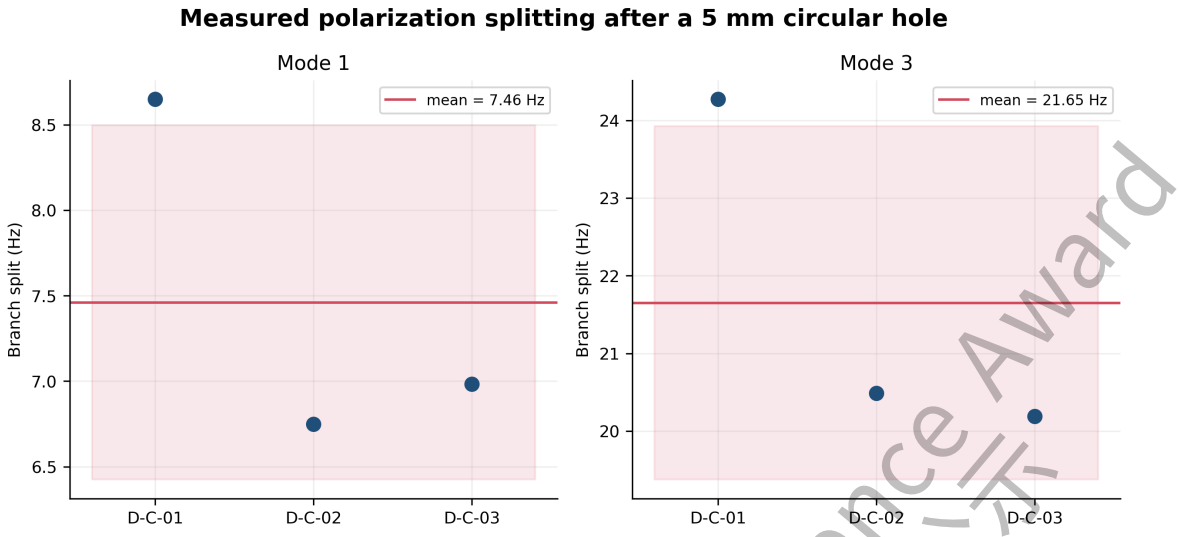


Figure 5. MEASURED Post-defect orientation-associated branch separation in three independently machined D-C tubes with midpoint 5 mm circular holes. Points are specimen values; horizontal lines and shaded spans show the mean and observed range. The repeatable positive separations in modes 1 and 3 exceed the corresponding control-derived noise by large factors; because matched intact two-orientation baselines were not retained, the full magnitude is not attributed uniquely to drilling.

Table 3. External group-level nominal-diameter-label validation. C specimens were excluded from the D-C calibration and GPR training. Internal GP standard deviations (approximately 0.005 mm) are omitted because they quantify interpolation of a deterministic physics curve, not total experimental uncertainty.

Specimen	Nominal label (mm)	GPR estimate (mm)	Absolute difference (mm)
C-3mm-01	3.00	3.21	0.21
C-3mm-02	3.00	2.68	0.32
C-5mm-01	5.00	4.61	0.39
C-5mm-02	5.00	4.74	0.26
C-8mm-01	8.00	7.38	0.62
C-8mm-02	8.00	7.21	0.79

was 7.7×10^{-6} in χ^2, confirming that equal minima are a symmetry of the declared three-frequency forward computation rather than an approximately close coincidence.

The estimator reports the quantity directly supported by the measurement: distance from the nearest end. Within that quotient space, the resolution is a few millimetres; in the original physical coordinate, the side is unsupported by the retained symmetric frequency features. Choosing a mirror after consulting the labelled position would not be an inference. Possible side cues in unmodelled waveform/setup asymmetries were not evaluated.

4.5 Held-out nominal-diameter label validation

MEASURED validated against SIMULATED physics training. The D-C-calibrated physics–GPR surrogate was evaluated on six real C campaign specimens that were excluded from this calibration. With no microphone-position offset applied, nominal-label MAE was 0.432 mm, median absolute difference was 0.35 mm, and $R^2 = 0.946$. Across the fixed measured-pilot offset scenarios, nominal-label MAE ranged from 0.355–0.432 mm. A direct physics-grid comparator produced nearly identical predictions, showing that the GPR emulates the calibrated curve rather than adding independent information.

All six nominal drill-diameter labels were correctly rank-ordered by class, but the two 8 mm specimens were underestimated by 0.62 and 0.79 mm. The no-offset 0.432 mm mean absolute difference is therefore

2026 Si Yuan High School Science Award
仅用于2026年度校内中学科学论文公示

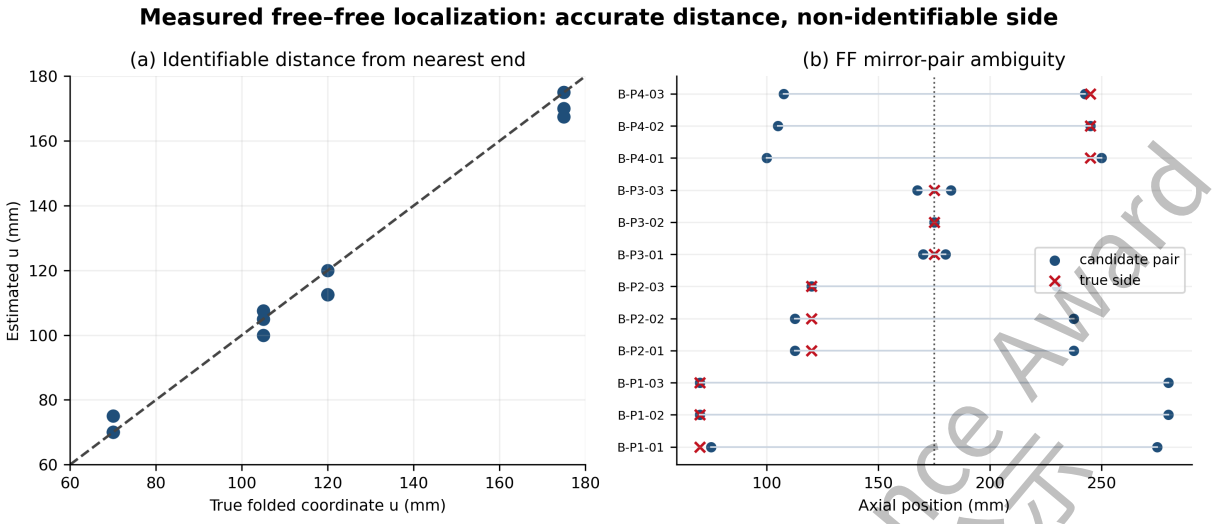


Figure 6. Measured Free–free localization. (a) Estimated versus true folded coordinate u, the quantity identifiable from a symmetric free–free spectrum. (b) Each observation admits the two physical candidates x and $350 - x$ with equal model status; red crosses show the labelled ground-truth side for evaluation only. No choice between the pair is supported by the retained FF frequency features under the symmetric model.

Table 4. Pilot Protocol-transfer sensitivity anchored to the six-recording microphone-position pilot. Each row reapplies one fixed offset vector to the existing campaign shifts and reruns the declared B localization and held-out C nominal-label pipelines. Mean±SD rows are sensitivity scenarios, not confidence limits or additional specimens.

Fixed offset scenario	Folded-location MAE (mm)	Maximum folded-location error (mm)	Nominal-label MAE (mm)
No offset	3.33	7.5	0.432
Pilot mean	3.33	10.0	0.376
Pilot mean −1 SD	3.54	10.0	0.398
Pilot mean +1 SD	3.33	10.0	0.355

a validation against nominal drill-diameter labels, not a metrology error against independently measured as-built diameters. This coherent bias is more informative than the tiny GPR interpolation variance: the limiting uncertainty is transfer from the D-C calibration to a different physical specimen group, not the numerical smoothness of the surrogate.

Measured calibration records with Simulated local propagation. Neither inverse branch was controlled by one calibration tube. Across the 9 exact leave-one-anchor refits of the general calibration, the unchanged B group retained folded-location MAE of 3.13–3.33 mm. Propagating the full-fit local coefficient covariance contributed approximately 0.82 mm standard deviation to the frozen location inverse. Across the 3 D-C-only refits, the still-held-out C group gave nominal-label MAE of 0.38–0.52 mm; the corresponding local coefficient-covariance contribution was approximately 0.050 mm. These are calibration-leverage checks, not new validations or total uncertainty estimates: they leave the coherent 8 mm bias and the elevated model residual untouched.

4.6 Hybrid Multi-BC localization: information gain and fragility

Hybrid The analysis fuses each specimen's real FF shift with pilot-constrained simulated CL/CR shifts. It is a conditional design-sensitivity study, not a measured fleet-level localization score. Under nominal pilot-constrained assumptions, the 3,000-draw-per-specimen analysis selected the correct side in 93.8% of draws. When the between-specimen calibration-residual standard deviation was added to both FF and simulated clamp variance components, correct-side probability fell to 85.5% and mirror-flip probability

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

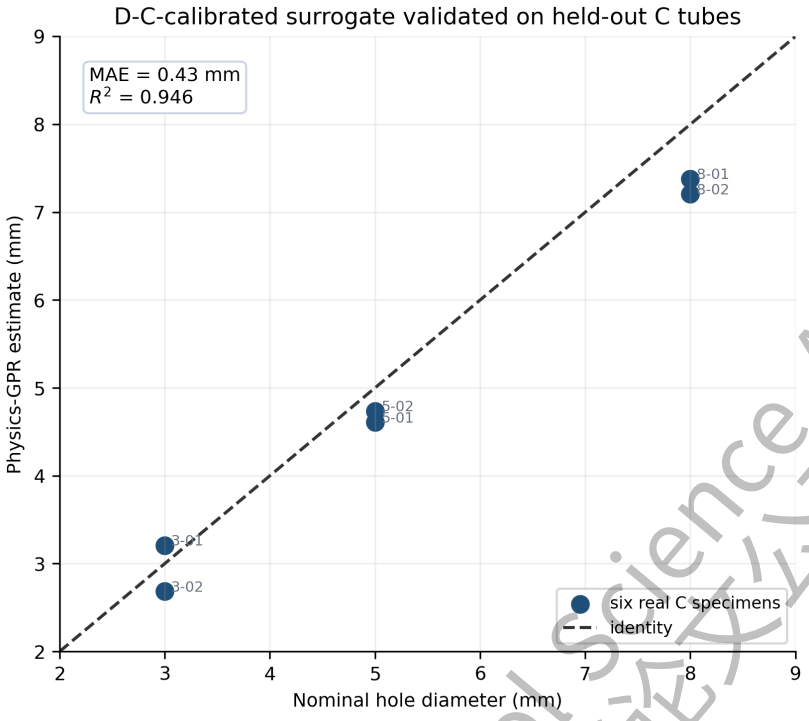


Figure 7. Measured held-out C-group nominal-diameter estimates from a Simulated D-C-calibrated physics–GPR surrogate under the no-offset scenario. The two 8 mm labels are systematically underpredicted, a visible model-transfer warning despite the high overall R^2; table 4 reports the measured-pilot fixed-offset sensitivity range.

rose to 14.5%. The correct-side conditional mean absolute error in that residual-inflated scenario was 3.95 mm.

The scenarios separate two failure modes. Within the correct basin, conditional error remains near $3.6 - -4.0$ mm; as clamp and model uncertainty increase, the wrong mirror basin is selected more often. The empirically anchored residual-inflated scenario therefore lowers the nominal correct-side result by 8.3 percentage points while leaving same-side precision near 4 mm. Thus the physical design claim is limited but useful: asymmetric boundary conditions can supply side information in the model, while real specimen-level CL/CR measurements are required to establish measured reliability.

4.7 Bayesian posteriors formalize the information limit

Measured The grid posterior makes the free–free symmetry visible as probability mass rather than only as equal profile scores. For every non-midpoint B specimen, the real-FF posterior assigns 0.50 of its marginal mass to each physical half of the tube, up to numerical rounding. This equality follows analytically from the declared symmetric forward map, grid, and priors; the empirical content is the measured folded-coordinate recoverability. The representative B-P1-01 posterior in figure 9 has equal peaks at the recorded station and its mirror. Thus a narrow estimate of folded distance is compatible with a two-valued physical-coordinate posterior: the unresolved side is an information limit of the symmetric measurement/model pair, not an optimization failure.

Hybrid In the lower panel of figure 9, the same real FF observation is combined with nominal CL and CR response vectors generated by the pilot-constrained finite-root model. For B-P1-01, this nominal-noise conditional posterior assigns 0.998 mass to the recorded half, whereas midpoint P3 specimens remain symmetric in physical side because the midpoint has no distinct mirror partner. It must be read together with the residual-inflated Monte Carlo result (85.5% correct-side probability and 14.5% mirror

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

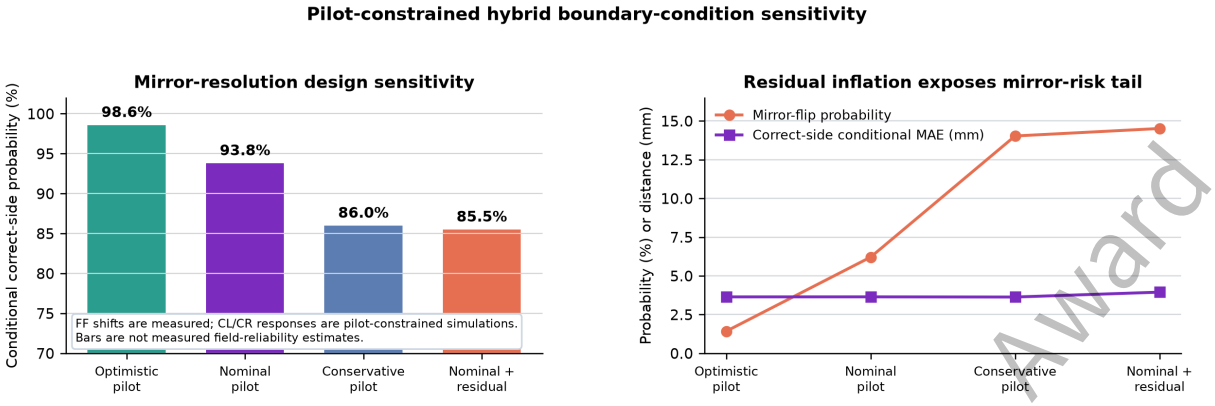


Figure 8. HYBRID Pilot-constrained Multi-BC design sensitivity. FF shifts are measured; CL/CR responses are simulated under pilot-constrained clamp and stated model-discrepancy assumptions. The bars are conditional correct-side probabilities, not measured field-reliability estimates. The right panel overlays mirror-flip probability and correct-side conditional MAE to show that basin-selection risk grows even when within-basin precision remains similar.

flips), rather than as an independent reliability figure. These posteriors are deliberately stronger than the measured FF posterior because they include an asymmetric intervention generated by the model; they demonstrate the side information that the proposed design could add, not a measured field posterior.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

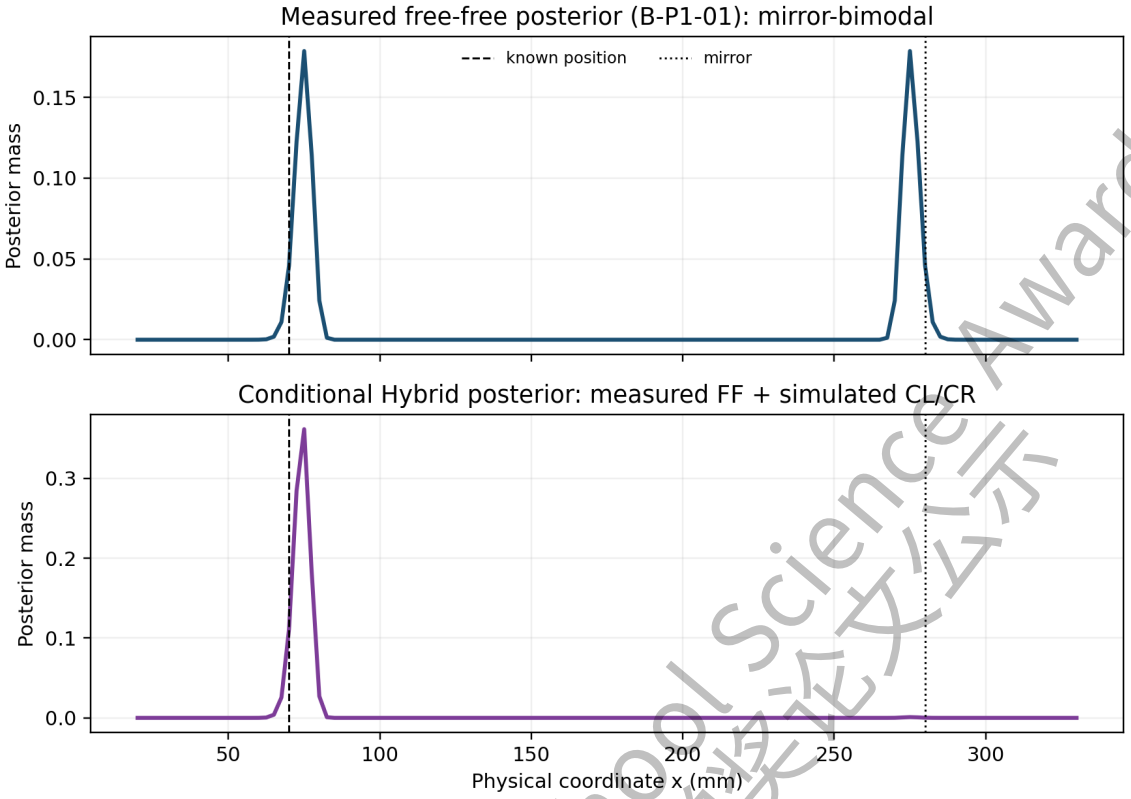


Figure 9. Bayesian localization posteriors for representative B-P1-01. **Top:** MEASURED posterior from the three real campaign free–free shifts. The equal peaks at the station and its mirror are imposed by symmetry. **Bottom:** HYBRID nominal-noise conditional posterior after adding nominal CL/CR shifts generated by the pilot-constrained finite-root model. The lower panel is a simulated experiment; it isolates the side information supplied by asymmetric boundary conditions and is not a measured CL/CR fleet result. Its scale is more optimistic than the residual-inflated Monte Carlo scenario.

4.8 Exploratory non-circular trials and controlled sensitivity analysis

MEASURED The non-circular after-state trials did not yield a stable, reproducible separation among the planned D-C, D-OT, D-OA, and D-SA fabrication families when only the three retained global bending modes were used under the tested fixture and acoustic configuration. Accordingly, we report neither an experimental four-shape accuracy nor an experimental three-class accuracy obtained by relabelling the same tubes. The result identifies a practical observability limit of the measurement system: fixture variability, environmental noise, microphone response/placement, and the intrinsically global nature of the retained features can all mask the small geometry-dependent differences expected between non-circular defects. The present experiment does not separate those contributors quantitatively.

SIMULATED To make that non-significant result informative, the shape sensitivity analysis was anchored to the three real D-C doublets; its fit had $\chi^2 = 36.47$ on 10 degrees of freedom. Under the nominal large-sample four-label scenario (0.10 mm machining SD and unit D-C-derived noise), modal-only LDA reached 91.9% mean test accuracy (5th–95th percentile 90.3–93.7%). Area-normalized anisotropy alone reached only 48.7%, showing that normalization removes much of the easy severity information. The area-informed vector $[\|\Delta \boldsymbol{f}_{\text{sens}}\|, s_1, s_3, q_1, q_3]$, with $q_i = s_i/A$, reached 98.5%; area A was used to form q_i but was not passed directly as a feature.

Those values describe a classifier trained on hundreds of simulated examples per class. At the planned 12-specimen scale, the paired taxonomy comparison in table 5 was lower and more variable. With equal class priors, the four nominal labels reached 82.3% modal-only mean accuracy and 90.9% with area-

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Table 5. SIMULATED Paired small-N taxonomy comparison on the same 300 sampled 12-specimen datasets. Each repeat retains three simulated specimens from each original fabrication family; labels are then remapped without deleting specimens. Scaler and shrinkage LDA are fitted inside each LOO fold with equal class priors. Values are design-sensitivity results, not experimental classification accuracy.

Taxonomy	Feature set	Accuracy (%)	Balanced (%)	Median /12	5th–95th /12	Class recall (%)
Four nominal labels	Modal-only	82.3	82.3	10	7–12	67.2–99.9
	Normalized anisotropy	40.6	40.6	5	2–8	28.8–51.7
	Area-informed	90.9	90.9	11	9–12	85.3–100.0
Three geometry families (circle / oval / slot)	Modal-only	86.5	89.2	11	8–12	78.6–100.0
	Normalized anisotropy	40.2	44.1	5	2–8	28.6–67.4
	Area-informed	95.9	96.4	12	10–12	94.4–100.0
Three spectral families (circle / transverse / axial-elongated)	Modal-only	49.3	50.4	6	4–8	43.1–62.0
	Normalized anisotropy	59.8	53.6	7	4–10	30.6–78.3
	Area-informed	64.0	62.7	8	5–11	54.7–68.0

informed features. Relabelling the identical draws into three geometry families—circle, pooled oval, and slot—raised mean accuracy to 86.5% and 95.9%, with balanced accuracy of 89.2% and 96.4%, respectively. The modal-only median improved from 10/12 to 11/12; the area-informed median improved from 11/12 to 12/12. Across aggregated predictions, the three-geometry class recalls were 78.6–100.0% for modal-only and 94.4–100.0% for area-informed features.

Because every taxonomy used the same sampled specimens, its gain can be read repeat by repeat. The three-geometry map gained 4.3 overall and 6.9 balanced-accuracy points over four labels in the modal-only comparison; it produced more correct classifications in 46.0% of repeats and fewer in 21.7%. With area-informed features, the gains were 5.0 and 5.5 points, with more correct in 49.3% of repeats and fewer in 13.0%. This is an improvement in the simulated design question, not a conversion of the measured null result into a successful classifier.

The alternative spectrally motivated collapse was not supported. Pooling D-OA and D-SA into an axial-elongated class gave only 49.3% overall and 50.4% balanced modal-only accuracy; even the area-informed result reached only 64.0% overall and 62.7% balanced accuracy. Area-normalized anisotropy alone remained weak for the four-label and three-geometry tasks (40.6% and 44.1% balanced accuracy). Thus, merging the pair that is nearly coincident in one normalized coordinate does not by itself create a coherent class across the unnormalized severity features.

The normalized feature geometry explains both the experimental difficulty and the taxonomy result. D-C and D-OT are only weakly separated in q_3 (Cohen's $d \approx -0.39$), while D-OA and D-SA are nearly coincident after area normalization ($d \approx -0.11$). D-OT is more distinct from D-OA and D-SA ($d \approx 1.8$–2.1). The three-geometry map is therefore retained as the preferred *coarse design target* because it has a clear contour-family meaning and improves paired small-sample stability and worst-class recall; the four fabrication labels remain the finer sensitivity test. However, the high three-geometry area-informed score is severity-assisted and requires independently known area to construct q_i. Neither taxonomy establishes a unique spectral record of contour, and both remain consistent with the non-significant

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

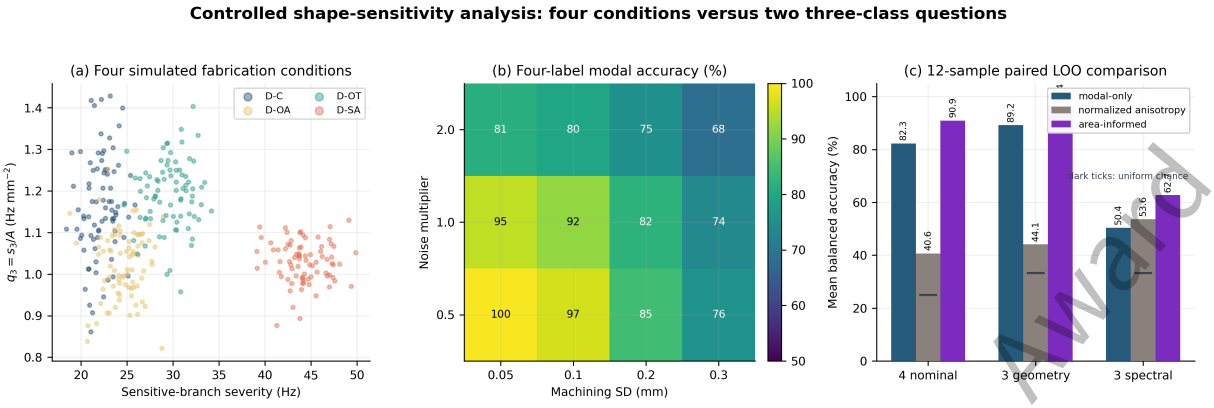


Figure 10. SIMULATED Controlled shape-sensitivity analysis, motivated by the non-significant non-circular experiment. (a) The four original fabrication conditions in severity–anisotropy space. (b) Four-label modal-only test accuracy falls as machining tolerance and noise grow; the severe corner reaches about 68%. (c) Equal-prior mean balanced accuracy from the same 300 paired 12-specimen LOO draws under four nominal labels, three geometry families (circle / oval / slot), and three spectrally motivated families (circle / transverse oval / axial-elongated). Dark horizontal ticks mark uniform chance. The three-geometry area-informed result is the most stable coarse design target, but normalized anisotropy remains weak and no bar is an experimental accuracy.

measured separation under the tested configuration.

4.9 Evidence-qualified summary

Table 6 summarizes the result statements and their evidential scope.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Table 6. Summary of what the study supports. The wording in the final column is part of the result: it prevents a simulation output from being re-labelled as experimental performance.

Question	Result	Evidence	Defensible interpretation
Defect present?	legacy 19/21 and campaign 18/18 above descriptive $D = 3$ rule; legacy 5/6 below in-sample; LOO range 17/21–21/21	Measured	protocol-bound laboratory screening; independent calibration absent
Orientation-associated separation?	mean 7.46 and 21.65 Hz in modes 1 and 3	Measured	repeatable post-defect directional separation; drilling contribution not uniquely isolated
Where is the defect?	folded-coordinate MAE 3.33–3.54 mm; calibration LOO 3.13–3.33 mm; equal FF posterior mass at mirrors	Measured+Pilot+Simulated	nearest-end distance robust to tested protocol and anchor perturbations; full-coordinate side remains an information limit
How large?	held-out nominal-label MAE 0.355–0.432 mm; calibration LOO 0.38–0.52 mm	Measured+Pilot+Simulated	one-station external-group proof of concept; anchor stability does not remove model-transfer bias
Can asymmetric BCs recover side?	93.8% nominal, 85.5% residual-inflated conditional correct-side probability	Hybrid	model-supported information intervention requiring real fleet validation
Which shape?	non-significant experiment; three-geometry balanced accuracy 89.2% modal-only and 96.4% area-informed in paired sensitivity study	Measured+Simulated	circle/oval/slot is a severity-assisted design target; four labels remain a finer stress test; no experimental shape accuracy

5 Discussion

5.1 The main result is an information budget

This study maps the information budget of one low-cost measurement. The same three resonances carry different kinds of information at different evidential strengths: their magnitude supports screening and severity; their cross-mode pattern supports position only modulo structural symmetry; the split between orthogonal polarizations reports cross-sectional anisotropy; and a contrast between asymmetric boundary conditions can, in principle, supply the missing side. Treating these as separate observables makes both positive and null outcomes useful: each result identifies either a supported quantity or the observation that must be added to recover it.

This information budget can be stated compactly:

$$\underbrace{\|\Delta \boldsymbol{f}\|}_{\text{presence / severity}} + \underbrace{\Delta \boldsymbol{f}/\|\Delta \boldsymbol{f}\|}_{\text{position modulo symmetry}} + \underbrace{\boldsymbol{f}_{0^\circ} - \boldsymbol{f}_{90^\circ}}_{\text{anisotropy}} + \underbrace{\boldsymbol{f}_{\mathrm{CL}} - \boldsymbol{f}_{\mathrm{CR}}}_{\text{side information}} . \tag{19}$$

The terms are schematic rather than statistically independent. Their value is conceptual: they prevent a magnitude effect from being called shape recognition, or a ground-truth mirror correction from being called localization.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公开

The result also refines the usual hierarchy of modal damage assessment [2]. Existence, location, type, and severity are not simply successive classifier labels. Each requires an observation operator with the relevant symmetry and sensitivity. A more complicated inverse algorithm cannot recover a variable that the measurement operator removes.

5.2 Detection: protocol discipline matters more than a perfect point estimate

The detection result demonstrates why protocol discipline is part of the inference. Under the declared drift-corrected rule, 19/21 legacy defects and 5/6 controls were on the expected side of $D = 3$ in-sample. When each control was held out and the remaining five redefined the center and spread, 1/6 held-out controls exceeded the rule and the legacy defect count ranged over 17/21–21/21. Campaign reporting remains 18/18 defects above the rule, but no campaign controls were recorded. The supported output is therefore a protocol-bound laboratory screening observation together with an explicit normalization-sensitivity range, rather than one apparently universal accuracy.

The transition from the legacy to campaign recordings is itself a result. The campaign did not merely collect cleaner versions of the same quantity. By controlling impact azimuth, it selected a consistent polarization branch. That changed the observation operator: signed shifts became comparable across specimens. This explains why the quadratic statistic could detect a legacy tube even when its signed vector was unsuitable for model inversion. In practical terms, mechanical measurement discipline did more than post-processing could have done.

A-01 adds useful information precisely because it is retained. Including it in the control variance makes the threshold less attractive but more credible, and shows that the same six controls both define the normalization and produce the apparent 5/6 in-sample control-negative count. The present data do not distinguish a rare acoustic/handling excursion from the tail of normal variation; that ambiguity specifies the next control design. A deployment study would interleave controls within each session and separate specimen, mounting, temperature, microphone, and trial variance, whereas the present analysis absorbs them into one between-session sigma.

5.3 Post-defect branch separation is both a signal and a warning

The D-C recordings establish repeatable post-defect orientation-associated separation in modes 1 and 3. This is physically consistent with experiments on imperfect cylindrical resonators, where local mass–stiffness imbalance separates paired natural frequencies [6]; that study supplies an analogy rather than a validation model for the present tube modes. Across three independently machined damaged tubes, the branch separations share the same sign and order and are much larger than the control-derived scale. Because matched intact two-orientation baselines were not retained, the experiment establishes the repeated separation while leaving its full partition between drilling and pre-existing anisotropy unresolved.

This observation has two consequences. First, the post-defect branch separation is an additional orientation-sensitive feature. Because both branches are measured on the same damaged specimen, their difference is less dependent on between-specimen baselines than a before–after shift. It therefore defines a falsifiable next experiment: matched intact orientations can partition the newly induced component and test whether the residual doublet measure supports baseline-light screening across hole sizes. Until that comparison exists, the doublet remains a supported hypothesis rather than a replacement for paired measurements.

Second, the resolved branch pair makes impact direction part of the measurand. A peak near 5.57 kHz is not fully specified by saying “mode 3” once the damaged tube supports distinguishable branches; the branch and excitation plane also matter. Earlier sign reversals were therefore not random model

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

failures. They were evidence that the protocol had not fixed the state being compared. This distinction generalizes: whenever a specimen supports separated modal branches, automated peak tracking should follow branches, not only nearest frequencies.

5.4 Free–free localization is accurate on a quotient space

The free–free result illustrates why inverse accuracy must be defined on the identifiable parameter space. The twelve real B specimens were localized to 3.33 mm MAE in $u = \min(x, L - x)$ in the no-offset computation. Propagating the measured microphone-pilot mean and mean±SD vectors through the unchanged inverse pipeline gave 3.33–3.54 mm MAE; the maximum specimen error moved from 7.5 mm to 10.0 mm in the nonzero scenarios. This is strong relative to the 350 mm tube length: the aggregate folded-coordinate conclusion survives the observed pilot-scale protocol perturbation, while the larger maximum warns that individual estimates are more fragile than the mean. These fixed-offset ranges are sensitivity scenarios, not confidence intervals. The result is consistent with the established use of changes in vibration characteristics for damage location [1]. But u is a quotient coordinate: points related by reflection are the same state under the FF observation.

The physically valid free–free output is the folded coordinate. Choosing whichever mirror is nearer the known position would use ground truth after inference and turn a sensitivity-curve calculation into an apparent full-coordinate estimator. The vanishing numerical χ^2 difference between mirror minima demonstrates that more training data using the same three frequency features and symmetric forward model cannot solve the side problem. The Bayesian grid calculation states the same fact probabilistically: for every non-midpoint B tube, the measured-FF posterior places half of its marginal mass at each physical side, up to numerical rounding. This equality is analytic under the declared symmetric model, grid, and priors; the empirical result is the measured folded-coordinate recoverability. The result is therefore an exactly two-valued inference, not an unusually uncertain numerical fit. Unmodelled side cues may exist elsewhere in the raw waveform or setup, but a data-driven model would need to distinguish such physics from label imbalance, filename convention, or another artifact outside the retained modal features.

The project therefore reaches a design principle consistent with recent theoretical stiffness-modulation work [4]: *when non-uniqueness is generated by a physical symmetry, alter the experiment before altering the optimizer*. CL and CR boundary conditions do exactly this by moving nodes and curvature fields differently relative to the two physical ends. The intervention here is boundary-condition contrast rather than active stiffness modulation or controlled deformation, but the shared mechanism is to change the forward operator so that the two states no longer share an image.

The hybrid result quantifies an intervention together with its failure mode. Under nominal pilot-constrained assumptions, the additional boundary responses yield 93.8% conditional correct-side probability. Adding the measured between-specimen calibration-residual component reduces this to 85.5%, with 14.5% mirror flips and 3.95 mm error conditional on remaining in the correct basin. The conditional Bayesian posterior reaches the same conclusion from another representation: nominal asymmetric CL/CR responses collapse the off-midpoint mirror pair, while the midpoint remains symmetric because it has no distinct side. Its representative 0.998 posterior mass is a nominal-noise quantity and must be read jointly with, rather than substituted for, the residual-inflated Monte Carlo figures. Near-constant correct-basin error alongside a rising flip probability shows that the design fails by choosing the mirror basin, not by gradually losing within-basin precision. This is an evidence-anchored conditional design prediction whose decisive test is real fleet-level CL/CR validation.

A decisive next experiment is small and concrete: record real before–after CL and CR data for a balanced subset containing both mirror positions, with torque or clamp force measured, and hold out at least one specimen at each position. That experiment would test the key side-information claim directly.

2026 年 Youth High School Science Award 项目评审专家公示

5.5 Sizing: cross-group validation matters more than the choice of regressor

The size result is one of the strongest quantitative outcomes because this branch deliberately re-fits its own D-C-only coefficients ($\gamma_k = 1.447$, $\gamma_m = 0.289$) instead of inheriting the nine-specimen general calibration. Six C tubes are then held out from coefficient calibration, feature scaling, GPR training, kernel selection, and model selection. The no-offset comparison produced 0.432 mm nominal-label MAE and $R^2 = 0.946$; across the measured-pilot fixed-offset scenarios, MAE remained within 0.355–0.432 mm. The direct physics grid and GPR gave nearly identical answers. GPR therefore contributes smooth, fast interpolation; it is not the source of the apparent physical accuracy, and the scenario range does not convert nominal drill labels into as-built metrology. This distinction is important in a small dataset, where a flexible regressor could otherwise hide circular validation.

The residual pattern also gives a more useful conclusion than the aggregate MAE. Both 8 mm holes were underpredicted, while 3 and 5 mm errors were smaller. The model is becoming biased as the defect leaves the regime represented by a 5 mm anchor. Possible reasons include nonlinear local shell deformation, geometry-dependent stiffness concentration, and the linear d/D knockdown. A larger diameter series should therefore be used to improve the forward model, not merely to refit a more flexible kernel.

The experiment validates equivalent circular-hole diameter at one axial station and two specimens per diameter. It does not validate volume loss for arbitrary location, partial-depth corrosion, multiple defects, or non-circular geometry. The small GP internal variance is irrelevant to these transfer errors and should never be presented as a physical confidence interval.

5.6 A coarser shape taxonomy clarifies the observation requirement

The non-circular branch produces two linked results: a measured observability boundary and a simulated design target. The completed non-circular after-state trials did not provide stable, repeatable four-family separation from the three retained global bending modes under the tested fixture and acoustic configuration. This leaves the machined geometries physically distinct while showing that the current operator compresses their differences below reproducible resolution. Fixture variability, environmental noise, microphone response and placement, and the global averaging of local contour may all contribute, although the experiment does not partition them. Relabelling the same recordings as circle, oval, and slot adds no observation, so neither taxonomy receives a numerical experimental accuracy.

The calibrated sensitivity analysis asks the complementary design question: if the D-C physics transfers, which classification target is most defensible under stated machining and acoustic uncertainty? On the same 300 paired 12-specimen draws and with equal class priors, four nominal labels gave 82.3% modal-only and 90.9% area-informed mean accuracy. The three-geometry map—circle, pooled oval, and slot—improved these to 86.5% and 95.9%; its corresponding balanced accuracies were 89.2% and 96.4%. The gain did not come from deleting difficult examples: all twelve original specimens remained in every remapped dataset. Aggregated class recall also became more even, rising from a 67.2–99.9% range under four-label modal-only classification to 78.6–100.0% for the three-geometry target.

The feature hierarchy explains why the strongest score remains conditional. At the planned small-N scale, area-normalized anisotropy alone gives only 40.6% balanced accuracy for four labels and 44.1% for three geometry families. Moreover, the apparently natural spectral collapse of D-OA and D-SA fails when the full features are restored: the circle / transverse oval / axial-elongated map reaches only 50.4% balanced modal-only and 62.7% area-informed accuracy. Global beam frequencies integrate local geometry into a small number of section deficits; several contours can share nearly the same normalized anisotropy while remaining far apart in severity. Area is not passed directly to the area-informed classifier, but it is required to construct $q_i = s_i/A$. The highest result therefore measures a severity-assisted coarse

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

contour question, not a unique spectral code for shape.

The physically useful decision is therefore hierarchical. Circle / oval / slot is the preferred coarse taxonomy for a future validation campaign because it has a clear contour-family meaning and gives the most stable paired small-sample result; the four original fabrication labels should remain as a nested, deliberately harder sensitivity test. The measured null and the taxonomy sensitivity together specify what the next experiment must observe. Beam-level modal data can support severity and coarse anisotropy, but detailed contour requires observables with shorter spatial scale or richer directional content. Candidate additions include local shell modes, guided-wave reflections, circumferentially distributed sensors, measured mode shapes, controlled scans over impact and microphone angle, and a fixture with measured, repeatable boundary conditions. These additions do not merely add samples; they interrogate geometric features that global bending frequencies average out.

5.7 Model discrepancy is a result, not a nuisance parameter

The circular calibration has $\chi^2/\mathrm{dof} \approx 4.44$ and the shape-anchor fit has $\chi^2/\mathrm{dof} \approx 3.65$. Both exceed the control-derived observation model. The discrepancy should not be eliminated by declaring a larger arbitrary sigma after fitting. It says that a one-dimensional knockdown model is missing repeatable structure.

The calibration-stability analysis separates that model-form problem from single-anchor leverage. Exact leave-one-anchor refits keep B folded-location MAE within 3.13–3.33 mm and held-out C nominal-label MAE within 0.38–0.52 mm. The local Laplace contributions of approximately 0.82 mm for location and 0.050 mm for nominal size are narrower than the protocol-transfer and coherent 8 mm effects. Thus, the main warning is not that one favorable calibration tube controls the answer; it is that a stable low-order calibration can still transfer with systematic bias. Local coefficient covariance and structural discrepancy answer different questions and should not be collapsed into one error bar.

Several mechanisms are plausible. A through-wall opening in a thin tube produces local shell deformation and stress concentration not represented by a uniform beam element. The effective stiffness deficit depends on branch orientation, defect contour, and possibly contact at the deburred edge. Because no intact two-orientation D-C baseline was retained, pre-existing extrusion ovality or seam anisotropy may also contribute to the measured separation and are absorbed into the fitted sensitive-branch coefficients. Nominal specimen dimensions vary, and the calibration treats density and Poisson ratio as fixed. The microphone pilot also shows that extracted mode-3 frequency changes slightly with observation position, which an eigenvalue-only model cannot explain. Finally, the clamp pilot constrains only one side and one repeated tube/setup, not a distribution of remounted roots.

For this reason, the sensitivity analyses explicitly add the observed between-specimen calibration-residual standard deviation to the localization and hybrid variance components. This is a transparent residual-inflation scenario, not a complete discrepancy-function model; it still cannot cover structural alternatives absent from the parametric model. A future uncertainty analysis should learn discrepancy across mode, position, and boundary condition only after acquiring real validation data under those conditions.

5.8 Validity boundaries and design consequences

The following conditions make the scope of each conclusion auditable. Each also identifies a control, measurement, or validation step required for transfer beyond the present experiment.

1. **The independent unit is the specimen, not the WAV.** Three impacts improve frequency estimation but do not triple the biological/statistical sample size. Headline denominators use tubes,

2026 S.T. Yau High School Science Award 仅用于2026年成果科学奖公示

and the 326-file count describes provenance rather than statistical power.

2. **Control coverage is small and protocol-mismatched.** Six controls support the legacy noise model; no controls were recorded in the campaign. Campaign specificity and full accuracy are undefined.
3. **Microphone positions differ between paired campaign stages.** The six-file pilot anchors fixed-offset sensitivity scenarios in which folded-location MAE is 3.33–3.54 mm and nominal-label MAE is 0.355–0.432 mm. These ranges show that the headline conclusions survive the tested perturbation, but the pilot does not span the full fleet or all sessions and the ranges are not confidence intervals.
4. **Only FF is measured across the main fleet.** Multi-BC localization is hybrid. Seven CL pilot impacts constrain one root condition; there is no real B-fleet CL/CR before–after validation and no corresponding CR pilot.
5. **The non-circular experiment was non-significant under the tested configuration.** Its global three-mode signatures did not yield stable, reproducible four-family separation, and relabelling the same records does not establish experimental three-class accuracy. The calibrated simulation recommends circle / oval / slot as a coarse future design target, but D-C anchors alone do not demonstrate transfer to ovals or slots.
6. **The beam model is deliberately low order.** Elevated calibration residuals show omitted local shell physics and specimen variability. The fitted γ coefficients should not be interpreted as universal material properties.
7. **Size validation is narrow.** Six midpoint tubes, two per diameter, test only circular through-holes from 3 to 8 mm. The coherent 8-mm bias warns against extrapolation.
8. **The grid discretizes position at** 2.5 mm**.** Same-side errors partly reflect this resolution. Finer grids cannot solve the mirror ambiguity.
9. **Mode 4 was unusable under the current recording and peak-tracking protocol.** All inferences rely on three modes; higher stable modes might add shape information if measured differently.
10. **A pristine baseline is required.** The measured pipeline compares each tube with its own intact state. Baseline-free field screening is not demonstrated.
11. **Machined holes are not natural damage.** Corrosion, partial-depth loss, cracks, multiple defects, coatings, supports, and operational loading may produce different acoustic and modal responses.
12. **The simulations are conditional.** Monte Carlo frequencies sample the specified parameter/noise model; they do not establish that real field errors follow those distributions.

5.9 Evidence-ordered claims and decisive next experiments

The study supports five claims ordered by evidence. First, a low-cost, orientation-controlled impact protocol provides a three-mode, control-normalized laboratory change signal whose normalization sensitivity can be quantified. Second, real D-C tubes exhibit repeatable post-defect orientation-associated global-bending separation, while a matched intact two-orientation baseline is required to partition its cause. Third, real free–free shifts locate a circular hole accurately on the folded coordinate; equal mirror minima and posterior mass identify physical side as missing information under the retained symmetric operator. Fourth, an independently calibrated physics–GPR map predicts held-out real nominal drill-diameter labels at one station, while the coherent 8 mm residual defines its transfer boundary. Fifth, the hybrid Multi-BC and simulated shape branches specify two testable interventions—asymmetric boundaries for side and circle / oval / slot as a coarse contour endpoint—without promoting either to measured fleet performance.

These claims imply a compact validation matrix rather than a larger classifier: real FF/CL/CR before–after recordings at mirror-paired positions; session-matched controls; measured clamp force; multiple

2026 S_T. Yau High School Science Award 用于2026正成科学论文公示

diameters at more than one station; and a repeated non-circular campaign with a controlled fixture, documented microphone geometry, and enough repeats to separate acoustic variability from geometry. That campaign should predeclare circle / oval / slot as the primary coarse endpoint while retaining the four fabrication labels as a nested stress test. If resources permit one additional sensing dimension, circumferentially resolved mode shapes or local guided-wave scattering should be prioritized because they address the specific information lost by global frequencies.

6 Conclusion

This study converts the broad goal of detecting, locating, sizing, and classifying defects from an impact sound into a more precise outcome: a map of what three resonances identify and which observation must be added when they do not.

MEASURED Three-mode free–free shifts provided a control-normalized laboratory change signal. Under the declared drift-corrected rule, 19/21 legacy defects and 5/6 controls were on the expected side of $D = 3$ in-sample; leave-one-control-out refits ranged over 17/21–21/21 legacy defects and placed 1/6 held-out controls above the rule. The orientation-controlled campaign placed 18/18 defects above the rule but contained no contemporaneous controls. Three damaged D-C tubes showed repeatable post-defect orientation-associated separation, averaging 7.46 Hz in mode 1 and 21.65 Hz in mode 3; the missing intact two-orientation baseline prevents unique causal attribution of the full separation to drilling. The repeated separation nonetheless shows why excitation orientation and branch identity must be controlled in the measurement protocol.

MEASURED The same real campaign shifts localized the folded coordinate $u = \min(x, L - x)$ with 3.33–3.54 mm mean absolute error across measured-pilot fixed-offset scenarios for the twelve position specimens. Under the symmetric three-frequency forward model, the two mirror candidates have equal predicted spectra and equal measured-FF Bayesian posterior mass, so physical side is absent from the retained free–free features. Possible side cues outside these features were not validated. Reporting the folded coordinate makes the successful estimate and the missing side equally explicit, and turns symmetry into the specification for a boundary-condition intervention.

MEASURED+SIMULATED An independently calibrated physics–GPR surrogate predicted six held-out real nominal drill-diameter labels with 0.355–0.432 mm mean absolute error across the measured-pilot fixed-offset scenarios ($R^2 = 0.946$ without an offset). Exact calibration leave-one-out refits retained B folded-location MAE of 3.13–3.33 mm and C nominal-label MAE of 0.38–0.52 mm, showing that neither result is controlled by a single calibration tube. The systematic underprediction of both 8 mm labels identifies the transfer boundary: anchor stability does not remove model bias beyond the represented severity regime, which is larger in scope than local coefficient covariance or GPR interpolation.

HYBRID Pilot-constrained simulated CL/CR measurements break the reflection symmetry in the forward model: nominal-noise conditional Bayesian posteriors collapse the off-midpoint mirror pair, while a midpoint tube remains symmetric. They selected the correct side in 93.8% of nominal Monte Carlo draws; adding the empirical calibration-residual variance reduced this to 85.5%, with 14.5% mirror flips. The displayed posterior illustrates nominal information gain, while the residual-inflated simulation defines the reliability boundary and identifies mirror-basin selection as the principal failure mode. This hybrid result supplies a quantitative target for real specimen-level Multi-BC validation.

MEASURED The non-circular after-state experiment did not yield stable, reproducible four-family separation from the three retained global modes under the tested fixture and acoustic configuration; relabelling the same records does not establish experimental three-class accuracy. SIMULATED On the same 300 paired 12-specimen LOO draws and with equal class priors, the four nominal labels gave 82.3%

2026 S.-T. Yau High School Science Award 应用于2026丘成桐中学科学奖公示

modal-only and 90.9% area-informed balanced accuracy. A coarser circle / oval / slot target improved these to 89.2% and 96.4%, whereas normalized anisotropy alone remained at 44.1% and the alternative axial-elongated spectral collapse was not supported. Circle / oval / slot is therefore the preferred coarse endpoint for a future experiment, with the four fabrication labels retained as a nested stress test. The simulated gain is severity-assisted and does not show that three resonances uniquely encode local contour.

The practical lesson is that inverse performance is governed first by experimental observability. Better statistics can quantify uncertainty and faster surrogates can accelerate a known map, but neither can restore information erased by symmetry or averaged out by an insufficiently controlled observation. For this tube, the next scientific advance is a controlled campaign with real asymmetric boundary conditions, matched controls, measured clamp force, documented microphone geometry, and repeated non-circular trials. That campaign would test the two remaining model-supported design claims directly—asymmetric-boundary side recovery and coarse circle / oval / slot discrimination—and determine whether the present information map can become a validated inspection protocol.

Author contributions

E.J. and H.L. contributed equally to experimental design, specimen preparation, data acquisition, analysis development, and manuscript preparation. C.Y. provided project supervision and critical review.

Data and analysis availability

The study materials include the recorded waveform archive, specimen metadata, analysis code, and the tables used to generate the reported results. Selected data and analysis materials are available in the supplementary package or from the authors on reasonable request, subject to file-size and school data-sharing requirements.

Use of AI tools

Tecent Yuanbao Hy3 was used on July 27 for 2 sessions as supplementary aids for language refinement and coding support. The students conducted the experimental work and, with their supervising teacher, reviewed the reported formulae, data analyses, citations, figures, and conclusions. The authors retain responsibility for the final submitted manuscript. The chat record has been provided as a supplement.

A Evidence scope

Table 7. Evidence categories used throughout the study.

Category	Basis	Interpretation
Measured	Paired laboratory recordings and deterministic analysis	A result computed from the recorded experiment.
Pilot	Supplementary laboratory recordings	A real protocol or boundary-condition constraint, not a fleet-level performance estimate.
Simulated	Calibrated physical/statistical sensitivity analysis	A controlled design or identifiability result under stated assumptions.

2026 S. T. Yau High School Science Award 仅用于2026日成阿甲学科学术论文公示

Category	Basis	Interpretation
Hybrid	Measured free–free data combined with pilot-constrained boundary modeling	A model-supported prediction that requires direct fleet validation.

B Supplementary materials

The supplementary materials document the specimen matrix, signal-processing procedure, sensitivity-analysis settings, and additional figures. They are intended to make the scope of each result clear: measured results are distinguished from pilot constraints, controlled simulations, and hybrid predictions.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

References

[1] Scott W. Doebling, Charles R. Farrar, Michael B. Prime, and David W. Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review. Technical Report LA-13070-MS, Los Alamos National Laboratory, 1996. URL https://www.osti.gov/biblio/249299.

[2] John J. Moughty and Joan R. Casas. A state of the art review of modal-based damage detection in bridges: Development, challenges, and solutions. *Applied Sciences*, 7(5):510, 2017. doi: 10.3390/app7050510. URL https://www.mdpi.com/2076-3417/7/5/510.

[3] Nicoleta Gillich, Cristian Tufisi, Christian Sacarea, Catalin V. Rusu, Gilbert-Rainer Gillich, Zeno-Iosif Praisach, and Mario Ardeljan. Beam damage assessment using natural frequency shift and machine learning. *Sensors*, 22(3):1118, 2022. doi: 10.3390/s22031118. URL https://www.mdpi.com/1424-8220/22/3/1118.

[4] Arka P. Reksowardojo and Ann C. Sychterz. Damage localization strategy in beams through stiffness modulation and controlled deformation. *Journal of Engineering Mechanics*, 152(10):04026058, 2026. doi: 10.1061/JENMDT.EMENG-8766. URL https://ascelibrary.org/doi/10.1061/JENMDT.EMENG-8766.

[5] J. A. Franco-Villafañe and R. A. Méndez-Sánchez. On the accuracy of the Timoshenko beam theory above the critical frequency: Best shear coefficient. *Journal of Mechanics*, 32(5):515–519, 2016. doi: 10.1017/jmech.2015.104. URL https://arxiv.org/abs/1405.4885.

[6] Yunfeng Tao, Yao Pan, Shilong Jin, Yonglei Jia, Kaiyong Yang, and Hui Luo. Trimming of imperfect cylindrical fused silica resonators by chemical etching. *Sensors*, 19(16):3596, 2019. doi: 10.3390/s19163596. URL https://www.mdpi.com/1424-8220/19/16/3596.

[7] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts, 2006. ISBN 0-262-18253-X. doi: 10.7551/mitpress/3206.001.0001. URL https://gaussianprocess.org/gpml/.

[8] Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta. Interval estimation for a binomial proportion. *Statistical Science*, 16(2):101–133, 2001. doi: 10.1214/ss/1009213286. URL https://projecteuclid.org/journals/statistical-science/volume-16/issue-2/Interval-Estimation-for-a-Binomial-Proportion/10.1214/ss/1009213286.full.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学生科学奖论文公示