Research Statement

Jiaming Li, Ph.D.

Assistant Teaching Professor

Acoustic Modeling | AI-Enabled Engineering Design | Embodied AI and Robotics | Acoustic-Aware World Models

Research Vision: From Acoustic Prediction Model to Acoustic-Aware Robot Intelligence

My research is guided by a simple question: how can we build models that make complex engineering systems more predictable, more designable, and ultimately more intelligent? My earlier work addressed this question in acoustics, where I developed interpretable analytical models for complex resonator's noise control frequency prediction and converted them into practical inverse-design tools. My future work extends the same philosophy to embodied AI: using physically meaningful, multimodal models to help robots understand action consequences, reason more safely, and execute more reliably in the real world.

This trajectory is also a natural reflection of my broader background. My training in mechanical engineering gave me a foundation in acoustics, modeling, and design optimization; my work at CGG strengthened my experience in large-scale signal processing and industrial data analysis; and my current academic and entrepreneurial work has expanded my focus toward AI-enabled engineering systems, robotics, and embodied intelligence. Across these settings, the common thread in my research has been the development of models that connect physical insight with computational intelligence.

Prior Research: Analytical Models for Resonator's Noise Control Frequency Prediction and Inverse Design

My prior research addressed a long-standing limitation in the prediction and design of Helmholtz resonators and related side-branch acoustic structures for low-frequency narrowband noise control. In engineering practice, these resonators are widely used for engines, intake and exhaust systems, and other acoustic systems where installation space is limited and the target noise frequency must be controlled precisely. However, the traditional Helmholtz model is mainly determined by cavity volume, neck area, and effective neck length. As a result, it often fails when the resonator geometry becomes more complex. In particular, it cannot reliably capture the influence of cavity contour, structural asymmetry, multiple openings, and coupled resonator layouts on the sound-reduction frequency. This leads to three practical problems: inaccurate sound reduction frequency prediction for nonstandard geometries, limited applicability to asymmetric and multi-neck structures, and heavy reliance on repeated finite-element trial-and-error during design. My prior research directly addressed these three limitations.

Contribution 1: I solved the problem that traditional Helmholtz models cannot capture cavity shape effects. Conventional models may predict nearly the same sound-reduction frequency for resonators with the same volume and neck parameters, even when their cavity contours are different. In reality, however, the internal geometry changes the standing-wave pressure distribution and therefore shifts the sound-reduction frequency. To address this limitation, I developed a geometry-aware analogy mass-spring system (AMSS) model. By discretizing the acoustic cavity according to standing-wave pressure isosurfaces and establishing an energy-conservation relationship between continuous mass-spring units and an equivalent oscillator, the model directly incorporates contour information into the frequency prediction. This allows the model to distinguish resonators that have similar global volume but different shapes. More importantly, it significantly improves accuracy: in representative cases, the conventional model produced prediction errors of 13%–19%, while the AMSS model reduced the error to within 3% (JASA, 2024; AIP Advances, 2024).

Contribution 2: I solved the problem that traditional models are not general enough for asymmetric and multi-neck resonators. The classical Helmholtz formulation is most reliable for simple, symmetric, single-neck configurations, but many practical resonators are asymmetric, multi-opening, or geometrically coupled. Without a fast and accurate analytical model, these structures are difficult to design systematically and often require extensive numerical iteration. To overcome this limitation, I extended the AMSS framework into an analogy serial-parallel mass-spring system (ASPMSS) model. In this model, the shared propagation region is represented as a serial subsystem, while the cavities on both sides of the neck are represented as parallel subsystems, making symmetric

resonators a special case of a more general asymmetric formulation. I also introduced an equivalent wide-neck method to incorporate multi-opening structures into the same framework. This produced a unified high-accuracy model for arbitrary-contour, symmetric, asymmetric, and multi-neck resonators, with validation errors below 2% in tested asymmetric configurations (JVA, 2025a,b).

Contribution 3: I solved the design inefficiency caused by inaccurate and limited analytical models. In conventional resonator design, engineers often follow a trial-and-error workflow: modify dimensions, rebuild the model, remesh, run finite-element simulation, evaluate the result, and then repeat the process until the target frequency is reached. This approach is time-consuming and provides limited design insight. Because the ASPMSS model is both accurate and generalizable, I converted it into a target-oriented inverse design method. Instead of starting from a guessed geometry, the designer can specify a target sound-reduction frequency or frequency band, define geometric and manufacturing constraints, and analytically solve for feasible resonator dimensions. Finite-element simulation is then used mainly as final independent verification rather than as the primary design engine. This method has enabled one-step design of straight side branches, asymmetric cavities, multi-neck annular cavities, and serial resonator systems, transforming resonator design from repeated simulation-based trial-and-error into an interpretable, reusable, and efficient analytical design workflow (Acta Acustica, 2026).

Together, these studies established a coherent research line in high-precision acoustic prediction and target-oriented resonator design. The central contribution is not only that the models are more accurate, but that they change how complex acoustic structures can be designed: from empirical adjustment and repeated simulation toward physically interpretable prediction, analytical inverse design, and efficient engineering decision-making. Figure 1 summarizes how this work addresses the three limitations of traditional resonator modeling and translates them into engineering impact.

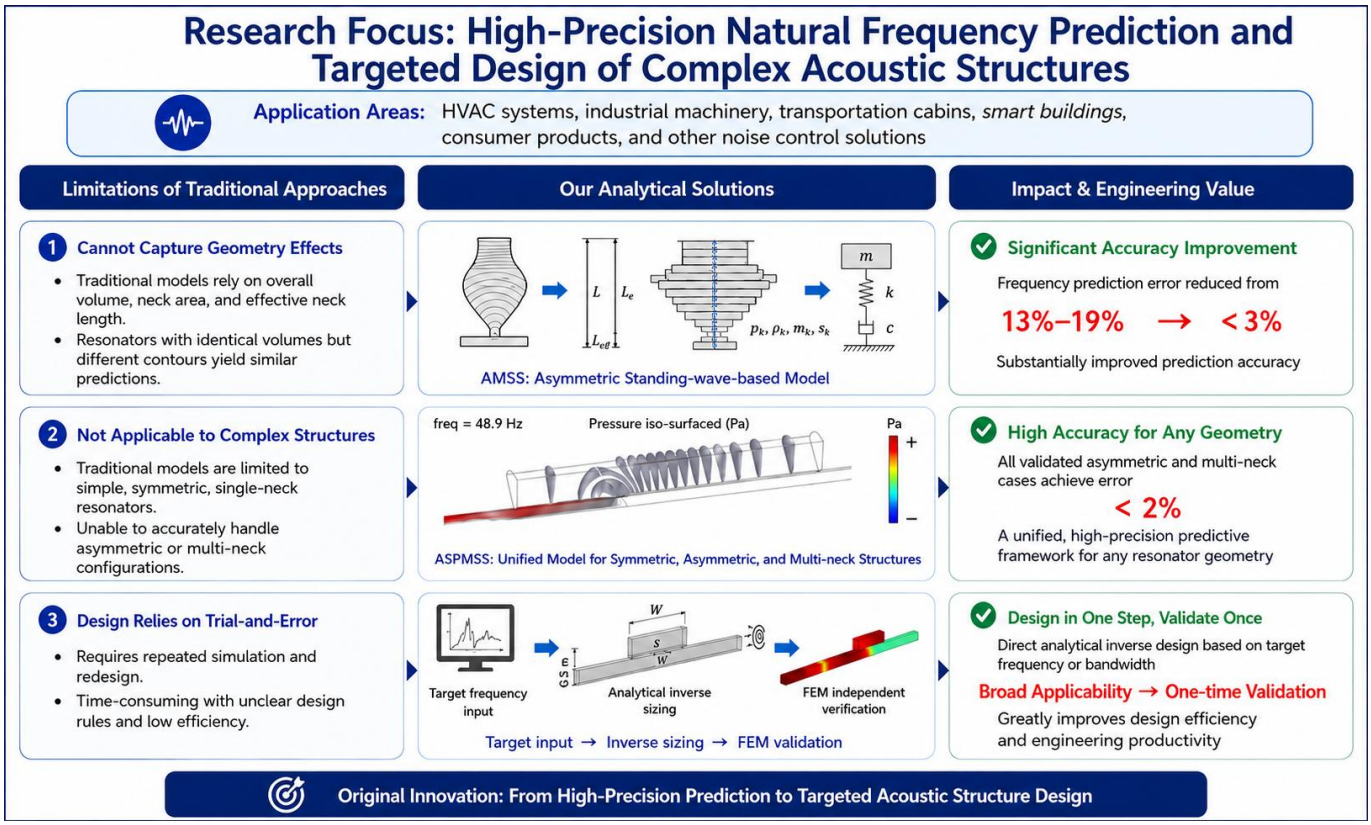

Figure 1. Summary of prior research contributions in acoustic resonator modeling and inverse design.

Future Research: Acoustic-Aware World Models for Robot Learning and Online Execution

The next stage of my research asks how the modeling and signal-intelligence expertise developed in acoustics can be used to improve embodied AI. Today's robot foundation models, especially vision-language-action (VLA) systems,

have made impressive progress in mapping images and language instructions to actions. Yet they still struggle in precisely the kinds of real-world situations where action consequences matter most: contact-rich interactions, subtle failures, uncertain observations, and safety-critical execution. In these settings, a robot should not rely on vision alone. It should also make use of acoustic and vibration cues, which often reveal slip, impact, abnormal contact, mechanical anomalies, and execution risk earlier than visual observation can.

I see three central challenges in current robot learning. The first is a ***data limitation***: real robot data are expensive, slow to collect, and difficult to scale, while simulation data often fails to capture the full physics of contact, friction, slip, and rare interactions. The second is an ***execution gap***: end-to-end VLA models perform well on routine tasks, but they are weaker at explicit action-consequence prediction, task decomposition, and risk-aware planning. The third is a ***deployment gap***: most systems rely primarily on vision, language, and state, leaving acoustic and vibration information largely underused even though these signals are highly informative for safety and anomaly detection.

To address these challenges, I propose an ***acoustic-aware world-model-driven framework for robot learning and execution***. The overall pipeline follows the same logic illustrated in Figure 2: first, build a multimodal dataset that combines real robot observations, real-world images and videos, motion-capture data, acoustic/vibration signals, and simulation or digital-twin data; second, train an acoustic-aware action-conditioned world model that predicts future states, outcomes, and acoustic events; third, use that world model to generate complementary experience and improve scenario coverage; fourth, train an acoustic-enhanced VLA model on the combined dataset; and finally, deploy the system in a dual-path execution architecture in which routine tasks are handled directly by the VLA executor, while complex or high-risk tasks are routed to the world model for imagination, decomposition, and planning before execution.

Figure 2 presents this framework graphically. The figure and the text are intended to align closely: the top row highlights the three bottlenecks, the middle row shows the full research pipeline from multimodal data to online execution, and the bottom row distills the core contributions of the study: efficient experience expansion, acoustic-aware consequence prediction and planning, and a feedback-driven loop for safer execution.

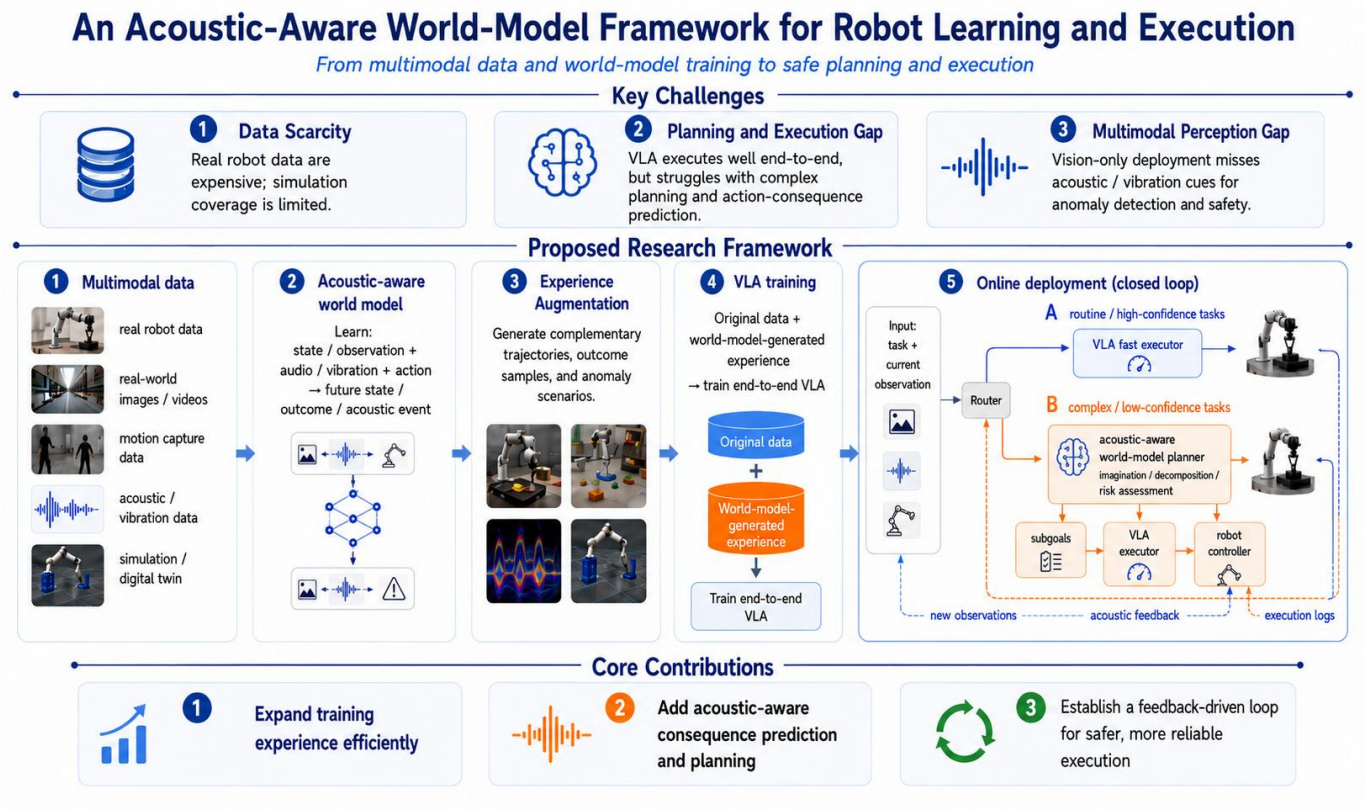

Figure 2. Proposed acoustic-aware world-model-driven framework for robot learning and online execution.

My future research program is organized around four connected objectives.

Objective 1: Build a multimodal robot learning dataset for contact-rich and anomaly-aware execution. This dataset will integrate robot camera observations, wrist-camera views, robot states, action trajectories, language instructions, acoustic/vibration signals, motion-capture data, and simulation or digital-twin data. The goal is not simply to collect more data, but to create a structured multimodal resource that captures the signals most relevant to manipulation, contact events, and failure diagnosis.

Objective 2: Train an acoustic-aware action-conditioned world model. This model will take current observations, robot state, actions, and acoustic/vibration cues as input and learn to predict future states, task outcomes, and acoustic events. Methodologically, I plan to use a latent-space world-model backbone that fuses vision, action, robot state, and audio/vibration representations. The model's role is not only to model dynamics, but also to provide interpretable action-consequence prediction under uncertainty.

Objective 3: Use the world model to expand training coverage and train an acoustic-enhanced VLA model. The world model will generate complementary trajectories, action-consequence samples, and difficult or anomaly-prone scenarios that are hard to obtain at scale from real robots alone. After quality filtering, these experiences will be combined with the original multimodal dataset to train an end-to-end acoustic-enhanced VLA model. In this way, the world model serves as a principled mechanism for improving data efficiency and scenario coverage rather than as a stand-alone simulator.

Objective 4: Deploy the system as a closed-loop robot executor with dual-path decision-making. At deployment time, routine or high-confidence tasks will be sent directly to the VLA executor for efficient action generation. Complex, low-confidence, or high-risk tasks will instead be routed to the acoustic-aware world model, which will perform imagination, action-consequence analysis, risk checking, and task decomposition. The resulting subgoals or action plans will then be executed by the VLA controller. New observations, acoustic feedback, and failure cases will be fed back into the training loop, forming a continuously improving closed-loop system.

Three-Year Plan, Expected Outcomes, and Research Development

The research timeline in Figure 3 matches the staged development of the program. In Year 1, I will establish the conceptual and data foundation through literature review, framework design, multimodal data construction, and acoustic/vibration representation learning. In Year 2, the focus will shift to world-model training, experience generation and filtering, and the training and post-training of the acoustic-enhanced VLA model. In Year 3, I will prioritize online deployment, planner integration, real-robot validation, feedback-driven system iteration, and broad dissemination through publications, software artifacts, and demonstrations.

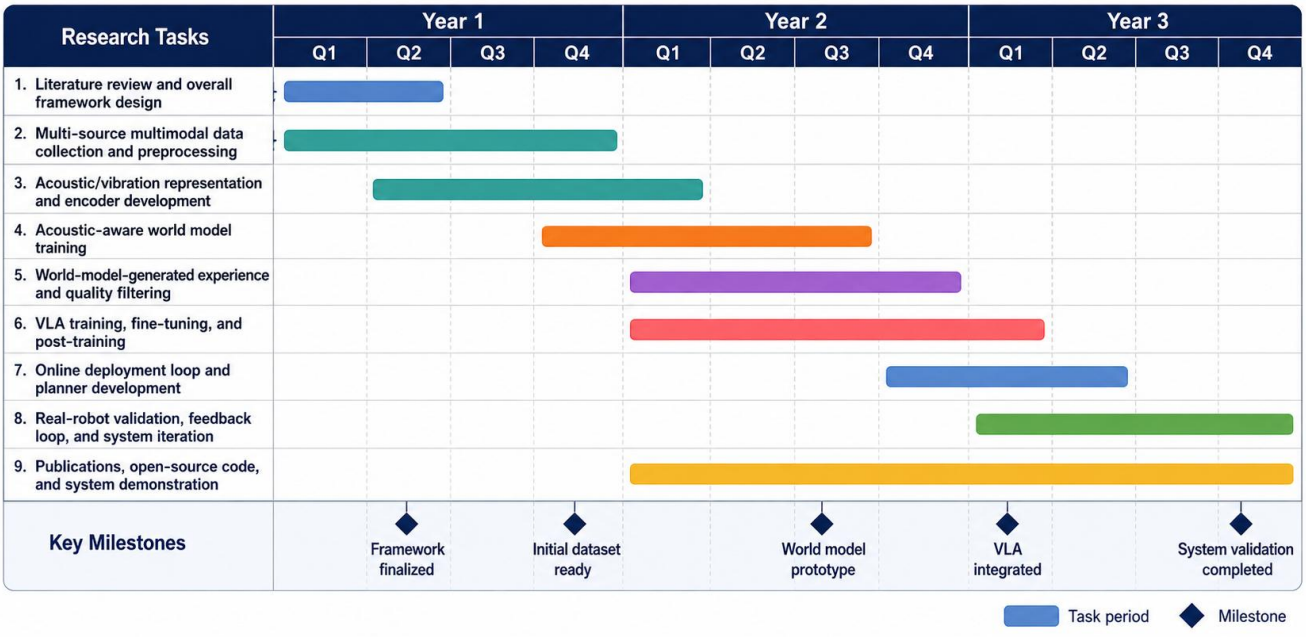

Figure 3. Three-year research timeline and milestones.

This program is designed to produce several concrete outcomes: a curated multimodal dataset for robot learning with acoustic/vibration cues; an acoustic-aware action-conditioned world model; a mixed training strategy that uses world-model-generated experience to improve VLA learning; a dual-path execution architecture for safer deployment; and a portfolio of publications, open-source tools, and external grant proposals. Over time, I expect this work to establish a distinctive research niche at the intersection of acoustics, multimodal AI, and embodied robotics.

This direction is also naturally collaborative. My ongoing collaboration with a Harvard Medical School/MGH team offers one pathway toward human-centered and medical applications of embodied AI, while my role as a co-founder of an embodied AI company provides a channel for industry-grounded problems and translational opportunities. Also, I plan to pursue external support through NSF and related interdisciplinary programs, complemented by collaborative and industry-linked projects.

Closing Statement

My research began with a practical engineering challenge: how to predict and design acoustic structures more accurately than traditional models allow. That work taught me how to turn physical insight into interpretable and useful models. I now aim to carry that same principle into embodied AI by building acoustic-aware world models and multimodal robot learning systems that help robots act more safely, more reliably, and more intelligently in the real world. My past and future research are a continuous effort to ***build models that make complex systems understandable, actionable, and ultimately more autonomous.***