Voice-Enabled Structural Health Monitoring via Large Language Models—A Cantilever Beam Case Study

GURU PRAKASH RAMAGURU
and VENKATA DILIP KUMAR PASUPULETI

ABSTRACT

Structural Health Monitoring (SHM) is critical in ensuring the safety and longevity of infrastructure such as bridges, pipelines, and buildings. Traditional SHM systems rely on sensor networks to capture vibrational and visual data, which are then analyzed using signal processing, statistical models, and machine learning to detect damage and predict failures. However, these systems often require expert interpretation and lack intuitive interfaces for real-time insights.

Recent advancements in artificial intelligence, particularly Large Language Models (LLMs), present new opportunities to enhance SHM workflows through natural language understanding and data-driven reasoning. In this paper, we introduce a novel LLM-powered voice agent that enables users to interact conversationally with structural sensor data. By integrating LLMs with time-series SHM data, our system can respond to voice queries, summarize experiment outcomes, detect anomalies, and generate visualizations on demand.

We validate our approach using experimental data from a cantilever beam subjected to multiple damage scenarios. These results highlight the potential of LLM-driven conversational interfaces to make SHM systems more accessible, interpretable, and actionable for both experts and non-specialists.

Keywords: Structural Health Monitoring, Large Language Models (LLMs), Anomaly Detection, Digital Twins, Experimental Study

Guru Prakash Ramaguru, Software Engineer, Coefficient, San Mateo, California, USA Email: guru@coefficient.io

Venkata Dilip Kumar Pasupuleti, Associate Professor, Mahindra University, Hyderabad, India Email: venkata.pasupuleti@mahindrauniversity.edu.in

INTRODUCTION

Structural Health Monitoring (SHM) plays a vital role in ensuring the safety, functionality, and longevity of critical infrastructure such as bridges, buildings, and pipelines[1]. By continuously monitoring structural behavior through sensors, typically accelerometers, strain gauges, or displacement transducers, SHM enables early detection of damage and deterioration, reducing the risk of catastrophic failures[2]. Over the past decade, SHM systems have evolved to collect increasingly rich, high-frequency data streams. These data are often multivariate, time-dependent, and sensitive to environmental changes, making their interpretation both computationally intensive and dependent on domain expertise[3]. As the scale and complexity of SHM data increase, there is a growing demand for systems that not only detect damage but also allow intuitive interaction with sensor information, empowering engineers and stakeholders to make informed decisions quickly.

Despite technological advancements in sensing and analytics, current SHM systems primarily rely on visual dashboards, scripted queries, and manual inspection of time-series plots or frequency spectra. These approaches require significant engineering knowledge and offer limited accessibility for non-experts. More critically, the interaction model is static and passive—users cannot easily ask questions like “What is the most recent anomaly on Beam 2?” or “Summarize today’s damage events.” The absence of natural language or voice-based interfaces creates a barrier to broader adoption and more effective use.

Since the introduction of BERT [4](Bidirectional Encoder Representations from Transformers) in 2018, large language models (LLMs) have grown rapidly, enhancing the capabilities of AI agents. BERT's bidirectional processing of text improved context understanding, setting the stage for more advanced models. OpenAI’s GPT-4o[5] further built on this foundation, enabling multimodal processing of text, images, and audio. Models like gpt-4o-mini-transcribe[6] have boosted transcription accuracy, while gpt-4o-mini-tts[7] enables lifelike text-to-speech synthesis. These innovations have demonstrated unprecedented abilities in understanding natural language, reasoning over structured and unstructured data, and even supporting real-time voice interaction. These capabilities present a unique opportunity to modernize how humans interact with SHM data.

In this work, we propose a novel integration of an LLM into the SHM workflow, enabling voice-based conversational interaction with structural data. Our system combines a time-series sensor database with an LLM-powered agent[8] (GPT-4o) capable of processing speech inputs, executing SQL queries and Python analysis scripts, and returning natural language responses either as text or synthesized speech. This approach allows users to ask high-level questions such as “Is the cantilever beam showing signs of damage?” or “Generate a plot comparing damaged and undamaged frequencies,” and receive immediate, human-like answers. To our knowledge, this is the first application of an AI language model with full voice-loop capability speech-to-text, LLM reasoning, and text-to-speech for SHM interaction.

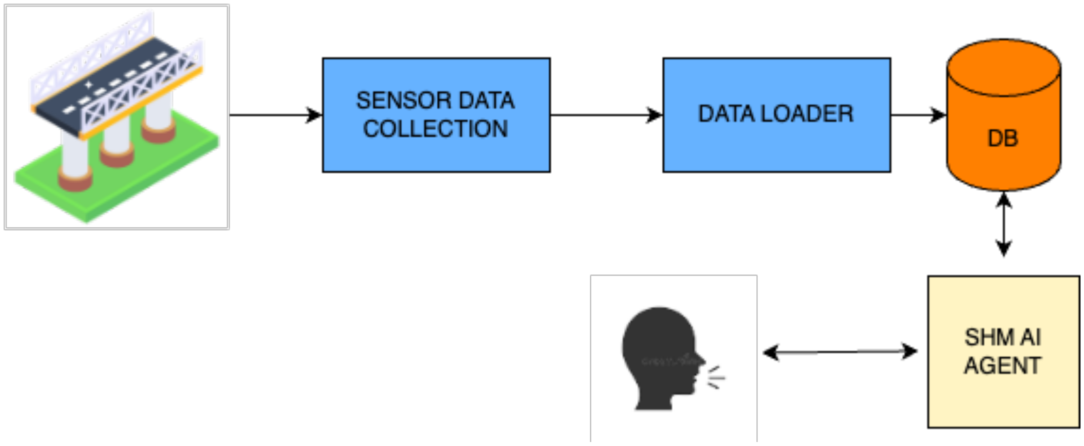

Figure 1. SHM Agent workflow

METHODOLGY

The system architecture for Structural Health Monitoring Agent follows a three-phase workflow that begins with Experimentation and Data Collection. The experimental setup and data were adapted from the paper [10]. A cantilever beam fitted with seven evenly spaced accelerometer sensors is subjected to various structural damage simulations. These experiments vary in terms of damage location, depth, and the type of excitation applied to the structure. Each experimental condition yields unique vibration response data, which is collected by the sensors and forwarded for further processing. This phase ensures comprehensive coverage of structural behavior under multiple failure scenarios, forming a robust dataset for analysis.

In the Data Loading and LLM-Based Voice Agent phases, the collected sensor data is systematically organized and stored. Specifically, the vibration readings and metadata are structured into two relational tables in a SQL database, one for experiment descriptors and the other for time-series sensor data, enabling efficient retrieval and traceability. The SHM AI Agent, powered by a Large Language Model (LLM), acts as a voice-based conversational interface that allows users to query and analyze the experimental data using natural language. It transcribes spoken questions, interprets them using the LLM, retrieves relevant results from the database, and responds with synthesized speech. This seamless integration of data storage and AI-driven interaction simplifies the exploration of complex SHM datasets for researchers and engineers. We use LangChain[9], a python-based framework for implementing the SHM Agent.

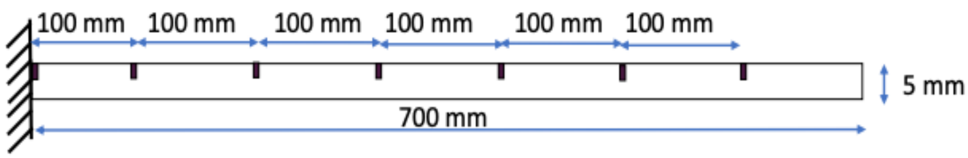

Figure 2: Cantilever experiment with 7 sensors

Sensor Data Collection

The data collection phase involved a systematic set of vibration experiments on a cantilever beam to simulate various structural damage scenarios. The beam was instrumented with seven accelerometer sensors, evenly spaced along its length. (Figure 2)

A total of 421 experiments were performed. Each experiment represents a unique combination of controlled variables, allowing for a broad exploration of damage effects under different physical conditions. For each experiment, approximately 10,000 time-series sensor readings were recorded across all sensors, yielding a rich dataset suitable for anomaly detection and pattern analysis. The experiment parameters are outlined in Table I.

Table I: Cantilever experiment parameters

Parameter	**Values / Description**
Damage Position	1mm, 2mm, 3mm from the hinge
Damage Depth	M1 to M7(1mm to 7mm depth)
Experiment type	Displacement, Single hit, Multi hit, Random hit

Data Loading

To enable efficient querying, traceability, and downstream analysis, the experimental metadata and time-series sensor readings were ingested into a structured SQL database. We used SQLite[11] as the database engine for its lightweight deployment and ease of integration with Python-based workflows. Two primary tables were created to represent the data

EXPERIMENTS

This table captures high-level metadata for each of the 421 experiments and serves as a foreign key reference for the time-series data. (Table II)

Table II: Experiments table structure

Column Name	Description
experiment_id	Unique identifier for the experiment
damage_position	1cm, 2cm, 3cm
damage_depth	M1 - M7
experiment_type	displacement, single_hit, multi_hit, random_hit
experiment_condition	RH 1-3-4 (random hits between 1 - 3 and then 3 - 4)

SENSOR READINGS

Each row in the sensor_readings table corresponds to a single timestamped set of vibration values across six axes. With ~10,000 readings per experiment, this results in over 4 million data points stored in the database. (Table III)

Table III: Sensor readings table structure

Column Name	Description
sensor_reading_id	Unique identifier for sensor reading
experiment_id	Id of the experiment
timestamp	Timestamp of the reading taken
experiment_type	displacement, single_hit, multi_hit, random_hit
sensor 1 – sensor 7	7 columns of readings from each of the sensors

SHM Agent Module

The voice interaction begins when a user speaks a query into a microphone. This audio input is captured in real-time and transcribed using **gpt-4o-transcribe**, a state-of-the-art speech recognition model optimized for varied acoustic environments. It delivers high transcription accuracy even in the presence of background noise or diverse accents, ensuring the user's intent is clearly captured in text form for downstream processing.

Once the spoken input is transcribed, the resulting text is processed by a custom Structural Health Monitoring (SHM) Agent built on **gpt-4o**. This agent understands the semantics of SHM-specific queries and dynamically generates **python scripts** to fulfill analytical tasks. Whether the request involves querying maximum sensor values, comparing vibration patterns, detecting anomalies, or visualizing trends, the agent executes the generated code using a secure backend tool, ensuring accurate and relevant results.

After processing the query and generating a result, the system formulates a natural language response, which may include summaries, numeric findings, or interpretive insights. This text is then converted back into speech using **gpt-4o-mini-tts**, a lightweight and expressive text-to-speech model. The result is a smooth, conversational response that enables users to receive information in real-time, creating a fluid and intuitive voice-based interface to interact with complex SHM data. (Figure 3)

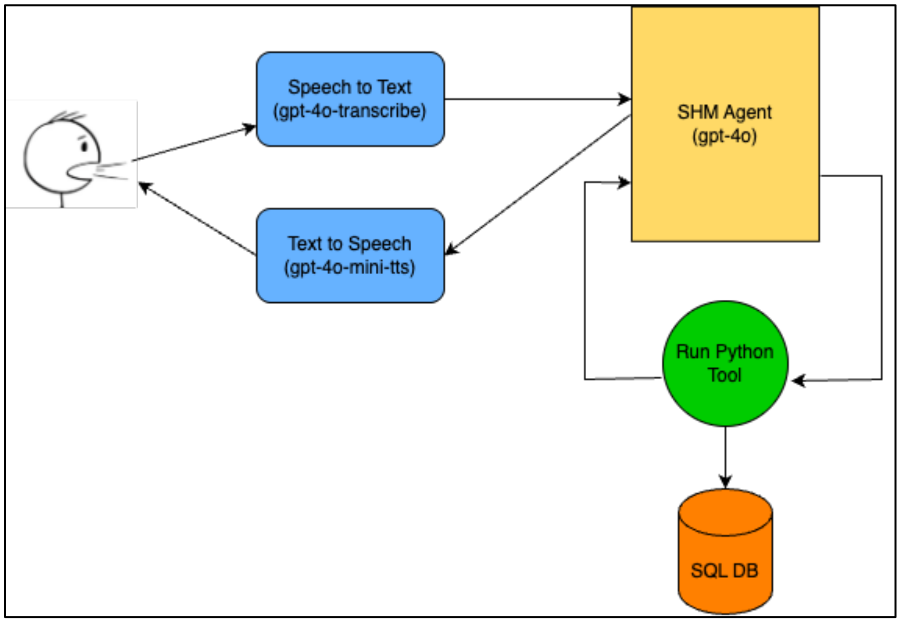

Figure 3: SHM Agent Architecture

Table IV: Query performance

Query Category	Accuracy	Avg Latency
Experiments Overview	10/10 questions	~6 sec
Single Experiment	9/10 questions	~8 sec
Pairwise Comparison	9/10 questions	~8 sec
Cross Experiment	8/10 questions	~12 sec

RESULTS

To evaluate the performance of the LLM-based voice agent, we designed a test using 40 voice queries divided into four categories, each reflecting common user interactions with the SHM system**.**

Experiments Overview queries assessed the agent's ability to summarize general dataset characteristics, such as "How many experiments were performed?" or "Visualize the number of experiments per damage position.", **Single Experiment** Queries focused on specific experiments, like identifying sensor readings above a threshold. **Pairwise Comparison** involved analyzing differences between two experiments, for example, comparing sensor 3 data between M2 and M5. Lastly, **Cross-Experiment Analysis** tested the agent's reasoning across multiple experiments. The performance is shown in Table IV.

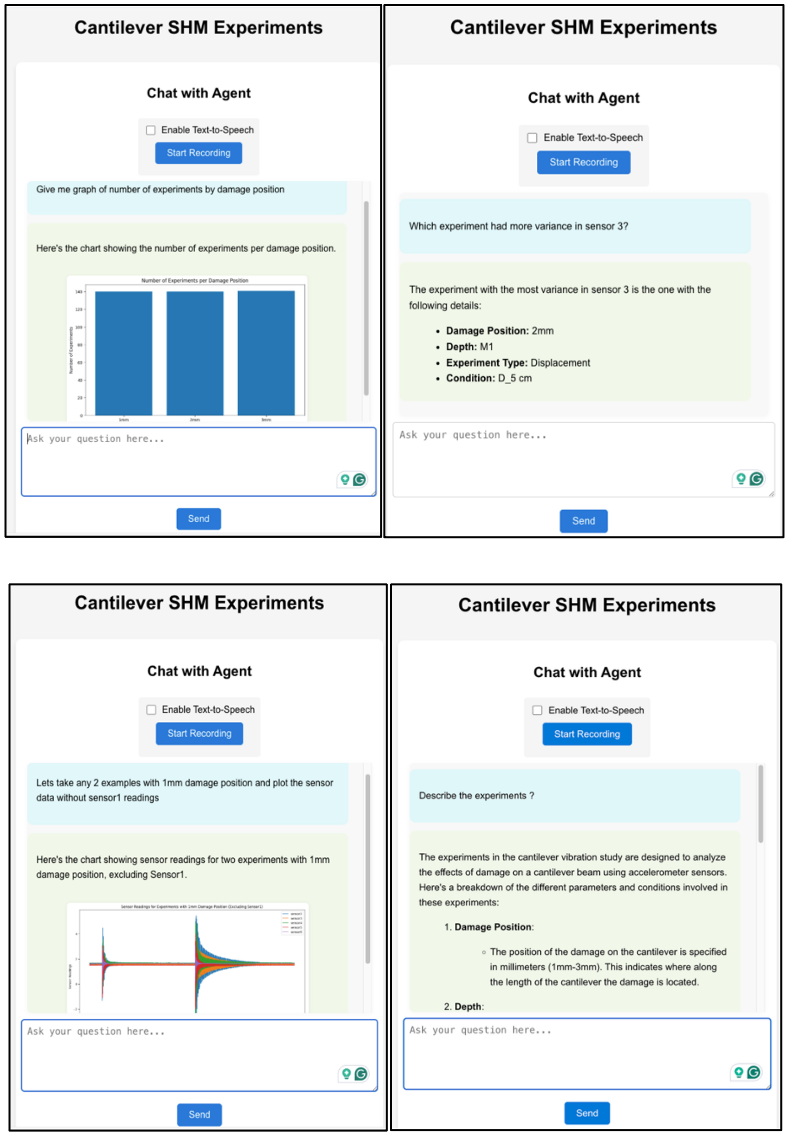

Figure 4: Example Agent Responses

Key Takeaways and Limitation

The LLM-based SHM agent demonstrated strong performance in accurately handling natural language queries across a range of experiment types. It effectively interpreted time-series sensor data, validating the role of large language models in Structural Health Monitoring contexts. A key strength of the system is its conversational interface, which lowers the barrier for users—especially those without technical expertise—to engage with SHM data. The agent maintained context well during multi-turn interactions and responded effectively to questions like "How many experiments were performed?", "Show a graph of sensor 3 responses for experiment M2", and "Compare average sensor readings between M1 and M3." These capabilities highlight the system's usability and accuracy in delivering relevant insights.

The work removes the need for domain-specific tools or SQL skills, allowing field technicians and engineers to obtain insights in real-time through natural conversation. It also sets the stage for AI-driven SHM assistants that can not only explain behaviors but also alert users to potential damage. However, limitations remain—such as latency during complex aggregation queries across 10+ experiments, or struggles with identifying anomalies without explicit definitions. For example, the agent performed poorly when asked "Which sensors show abnormal behavior over time?" or "Find all cases of unexpected drops in sensor 4." These shortcomings are tied to current LLM context length limits (128k tokens) and a lack of domain-specific anomaly reasoning.

Future Work

Future work will focus on scaling the SHM voice agent to more complex, real-world structures like bridges, pipelines, and aircraft components, where higher sensor density and intricate vibration patterns pose new challenges. We aim to enhance the system's ability to reason over numerical data by integrating domain-specific signal knowledge and exploring hybrid models that combine LLMs with physics-based simulations. Deployment in continuous monitoring environments will enable field technicians to interact with the agent for real-time diagnostics and trend detection. Additionally, robustness can be improved through the incorporation of complementary tools like computer vision for visual inspections and knowledge graphs for enriched contextual reasoning. Ultimately, this research lays the foundation for AI-powered infrastructure monitoring systems where LLMs act as reliable, conversational co-pilots in safety-critical settings.

REFERENCES

1. Katam, R., Pasupuleti, V.D.K. & Kalapatapu, P. A review on structural health monitoring: past to present. Innov. Infrastruct. Solut. 8, 248 (2023). https://doi.org/10.1007/s41062-023-01217-3
2. Kaya, Y., Safak, E. (2013). Real-Time Structural Health Monitoring and Damage Detection. In: Catbas, F., Pakzad, S., Racic, V., Pavic, A., Reynolds, P. (eds) Topics in Dynamics of Civil Structures, Volume 4. Conference Proceedings of the Society for Experimental Mechanics Series. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6555-3_2
3. Liu Z. Smart Sensors for Structural Health Monitoring and Nondestructive Evaluation. Sensors (Basel). 2024 Jan 17;24(2):603. doi: 10.3390/s24020603. PMID: 38257695; PMCID: PMC10819668.
4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), pp. 4171–4186. https://arxiv.org/abs/1810.04805
5. OpenAI. (2024). GPT-4o. https://openai.com/index/gpt-4o
6. OpenAI. (2024). gpt-4o-mini-transcribe. https://platform.openai.com/docs/models/gpt-4o-mini-transcribe
7. OpenAI. (2024). gpt-4o-mini-tts. https://platform.openai.com/docs/models/gpt-4o-mini-tts
8. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing Reasoning and Acting in Language Models. arXiv preprint arXiv:2210.03629.
9. Chase, H. (2022). LangChain: A Framework for Developing Applications with LLMs. GitHub Repository.

10. Katam, R., Pasupuleti, V.D.K. & Kalapatapu, P. SVM-assisted damage identification in cantilever steel beam using vibration-based method. Innov. Infrastruct. Solut. 9, 149 (2024). https://doi.org/10.1007/s41062-024-01459-9
11. Hipp, W. SQLite: A Relational Database Management System. 2000. https://www.sqlite.org/