

Scene2Hap: Generating Scene-Wide Haptics for VR from Scene Context with Multimodal LLMs

Arata Jingu*
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
jingu@cs.uni-saarland.de

Easa AliAbbasi*
Sensorimotor Interaction Group
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
easa.aliabbasi@mpi-inf.mpg.de

Sara Safaee
Sensorimotor Interaction Group
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
sara.safaee@uni-saarland.de

Paul Strohmeier
Sensorimotor Interaction
Max Planck Institute for Informatics,
Saarland Informatics Campus
Saarbrücken, Germany
paul.strohmeier@mpi-inf.mpg.de

Jürgen Steimle
Saarland University, Saarland
Informatics Campus
Saarbrücken, Germany
steimle@cs.uni-saarland.de

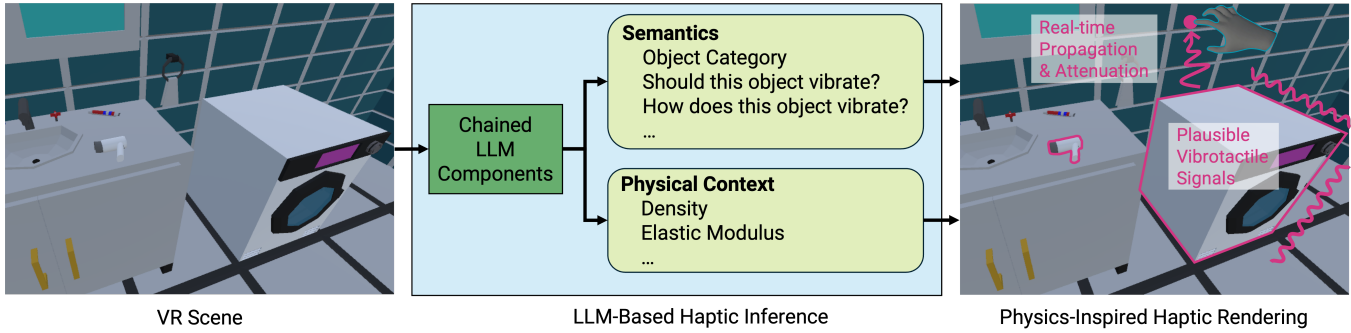


Figure 1: Scene2Hap is an LLM-centered system that automatically generates vibrotactile feedback for full VR scenes. It combines LLM-Based Haptic Inference, which extracts semantics and physical context of objects from multimodal scene data, with Physics-Inspired Haptic Rendering, which models how vibrations propagate and attenuate across objects in the scene, based on their LLM-inferred properties and physical context.

Abstract

Haptic feedback contributes to immersive virtual reality (VR) experiences. However, designing such feedback at scale for all objects within a VR scene remains time-consuming. We present Scene2Hap, an LLM-centered system that automatically designs object-level vibrotactile feedback for entire VR scenes based on the objects' semantic attributes and physical context. Scene2Hap employs a multimodal large language model to estimate each object's semantics and physical context, including its material properties and vibration behavior, from multimodal information in the VR scene. These estimated attributes are then used to generate or retrieve audio signals,

subsequently converted into plausible vibrotactile signals. For more realistic spatial haptic rendering, Scene2Hap estimates vibration propagation and attenuation from vibration sources to neighboring objects, considering the estimated material properties and spatial relationships of virtual objects in the scene. Three user studies confirm that Scene2Hap successfully estimates the vibration-related semantics and physical context of VR scenes and produces realistic vibrotactile signals.

CCS Concepts

• **Human-centered computing** → **Haptic devices.**

Keywords

Haptics; vibrotactile; generative; large language model; virtual reality; context.

ACM Reference Format:

Arata Jingu, Easa AliAbbasi, Sara Safaee, Paul Strohmeier, and Jürgen Steimle. 2026. Scene2Hap: Generating Scene-Wide Haptics for VR from

*Both co-first authors contributed equally.



This work is licensed under a Creative Commons Attribution 4.0 International License. CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04

<https://doi.org/10.1145/3772318.3791297>

Scene Context with Multimodal LLMs. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3772318.3791297>

1 Introduction

Designing 3D virtual worlds can be a tedious and time-consuming process, considering the number and diversity of objects present in realistic virtual reality (VR) scenes. To enable VR designers to generate these 3D virtual worlds in a scalable way, recent approaches have proposed using artificial intelligence (AI) or large language models (LLM) to automatically design their visuals, audios, or behaviors for a full scene comprising multiple virtual objects [20, 33, 72, 76].

However, designing the *haptic* properties of VR scenes remains challenging. Researchers have proposed generative machine learning models to design haptic signals from manually formulated text prompts or from images, for instance, with generative adversarial networks [83, 84] or LLMs [49, 77]. While these studies provide valuable insights regarding the automatic generation of haptic signals, they do not encompass two aspects essential to supporting scene-wide haptics: Firstly, they do not leverage the full semantic information of objects present in the VR scene. For example, a pot in a kitchen scene might not vibrate if found in a cupboard, yet it might vibrate intensely when boiling water on a stove. Secondly, they do not consider the physical context of objects and the relationships between multiple objects in the scene. For example, if a smartphone buzzes on a table, the actual vibration felt by the user depends on where the user touches the table and on the table's material properties; vibrations attenuate more quickly on a leather table than on a glass table. We believe that understanding object semantics and physical scene context is crucial for advancing haptic design in VR.

To overcome these limitations, we propose **Scene2Hap**, an LLM-centered system that automatically designs object-level vibrotactile feedback of an entire VR scene based on the objects' semantic attributes and physical contexts. In this work, we specifically focus on generating vibrotactile signals – the most frequently used form of haptic feedback in VR – that are triggered by active sources in the VR environment, such as machines or vibrating objects. For a given VR scene, Scene2Hap leverages a multimodal LLM to automatically estimate each object's semantics (e.g., whether and how the object vibrates) and material properties (e.g., density). It queries the LLM using the object's multimodal information present in the scene (e.g., images, name). We call this process *LLM-Based Haptic Inference*. The inferred object properties are used to create a plausible audio signal, which is then used as a vibrotactile signal after passing through a low-pass filter with a cutoff frequency of 250 Hz. Scene2Hap furthermore calculates a realistic vibrotactile signal, felt at the specific point the user touches in the scene, by considering the object's physical context: neighboring objects and the propagation of vibration across objects depending on their LLM-estimated material properties. Rather than assigning fixed vibration signals, Scene2Hap modulates them in real time, based on the user's touch location and the material properties inferred by the LLM. We call this process *Physics-Inspired Haptic Rendering*. The system delivers independent vibration feedback to each hand through handheld vibrotactile devices.

Results from three studies revealed that Scene2Hap (1) could successfully infer the semantics and physical contexts of objects in VR scenes; (2) significantly contributed to providing immersive VR haptic experiences by improving the sense of materiality and spatial awareness with vibration propagation and attenuation in the scene; and (3) successfully enhanced the overall user experience when the user interacts in a full VR scene designed with our end-to-end pipeline.

In summary, Scene2Hap proposes to consider scene-wide context for haptic rendering in VR, providing LLM-based haptic inference and physics-inspired haptic rendering in a novel architecture. These contributions position Scene2Hap as a new direction for scalable haptic design – one that links semantic inference with physics-inspired modeling to generate adaptive and realistic feedback for full VR scenes. We believe this hybrid approach can help make rich, real-time haptics a default capability in future virtual and mixed reality experiences. The main contributions of this work are:

- A novel system architecture, Scene2Hap, that automatically designs object-level vibrotactile feedback for full VR scenes by combining semantic inference and physics-inspired modeling.
- LLM-based haptic inference, which estimates semantic and material properties of virtual objects from automatically extracted multimodal scene data.
- Physics-inspired haptic rendering, which modulates vibrotactile feedback in real time based on inferred material properties, spatial arrangement, and user contact position.
- Empirical validation in three user studies showing that Scene2Hap successfully estimates the vibration-related semantics and physical context of VR scenes and produces realistic vibrotactile signals.

2 Related Work

Our work builds on the intersection of three areas: haptic design for VR scenes, machine learning-based haptic generation, and propagation of haptic signals in physical context.

2.1 Haptic Design for VR Scenes

Designing haptic attributes for VR experiences is a very complex task due to the need for extensive knowledge [42, 65, 70]. Various GUI-based haptic design tools have been proposed to make haptic design easier. They provide rich functions with the VR haptic designers, such as creating a new haptic signal from low-level parameters (e.g., amplitude, frequency, and spatiotemporal movement) [23, 36, 55, 67, 69, 85], editing existing haptic signals [38, 66], triggering a haptic signal in response to a specific event [66], and building a library of haptic signals [37]. For further rapid prototyping of haptic signals for VR scenes, in-situ VR haptic design methods based on designer-defined cues have also been proposed. These allow for designing and testing haptic signals directly in a VR scene, such as designing temporal signals for a hand-held haptic feedback device through the designer's vocalization [22] and designing spatiotemporal haptic signals for the whole hand based on the designer's spatial input and the hand's posture [78].

While these approaches have made haptic design more accessible, they still rely on manual effort to create and assign haptic signals to

individual objects. This becomes impractical in complex scenes with many interactive elements. Scene2Hap addresses this limitation by automating object-level haptic design using LLM-based inference, enabling scalable haptic generation across entire VR scenes without requiring low-level parameter tuning or manual signal authoring.

2.2 Machine Learning-Based Haptic Generation

To reduce the manual labor of haptic design, some recent works have proposed the automatic generation of haptic rendering signals using machine learning (ML) algorithms. Typical approaches have adopted generative adversarial networks to generate texture vibrations from image textures or material attributes [10, 13–16, 48, 83, 84]. Heravi et al. proposed an ML architecture to generate texture signals in real-time based on the user actions (force, speed) [29]. Faruqi et al. adopted a variational autoencoder to generate physical texture designs for 3D-printed objects [26].

Some very recent works started to leverage emerging LLMs to generate haptic signals from more free-form inputs, such as generating temporal vibrotactile signals from text prompts [49, 56, 77], generating spatiotemporal tactile patterns for gesture or emotion input [64, 73], and rendering appropriate thermal feedback based on a video context [57]. Conversely, LLMs have also been employed to interpret vibration signals into a textual description [31, 32].

Although ML and LLM-based approaches have enabled automatic haptic generation from images or text prompts, they typically operate outside the context of full VR scenes and require manual inputs for each object. As a result, they fail to capture how object semantics are shaped by scene context or how objects interact physically. Scene2Hap overcomes these issues by extracting structured, multimodal data from the scene and using chained LLM components to infer both semantic and physical object attributes in context, allowing for more automatic and context-aware haptic design.

2.3 Propagation of Haptic Signals on Surfaces

While the approaches discussed so far automate the generation of individual haptic signals, they treat objects in isolation and ignore how physical relationships between objects affect tactile perception. In particular, they do not account for how vibrations propagate across surfaces – an important factor for creating spatially coherent haptic feedback in VR. We next review work on vibration propagation to highlight this gap.

In a real-world scene, the vibrations generated by an object propagate through surfaces, a fact that is underestimated when designing haptic feedback for VR scenes. To the best of our knowledge, the impact of this vibration propagation on tactile perception has not been investigated yet for interactive VR experiences. However, studies in the sensory substitution domain showed increased spatial awareness among participants when representing locations of remote objects using vibrotactile signals [30, 40, 86]. Hence, we believe that considering the propagation and attenuation of vibrotactile signals while touching tactile surfaces in VR can influence multiple aspects of tactile perception.

The material's mechanical properties, such as its density, elasticity, and structural composition, can influence the speed and intensity of this propagation [2, 19, 39, 50, 62, 63, 80]. For instance,

vibrations propagate better through rigid and dense materials (e.g., metals) than soft and porous materials (e.g., rubber). To analyze vibration propagation, different structures are categorized based on their geometry and how they deform under loads [62, 79]. The most commonly investigated structures include plates [12, 80], strings [3, 17], bars [9, 61], shafts [82], membranes [62], shells [47], and beams [5, 25]. Propagation in each of these types is described by a different analytical model. Furthermore, the analytical solution can vary depending on the boundary conditions, e.g., whether the structure is free, simply supported, or fixed at its ends [62, 80]. To achieve the real-time behavior required for VR experiences, we simplify these mathematical and geometrical complexities and focus our approach on plates. We chose plates due to their vast availability in everyday home and office appliances. The goal is to provide a generic approach applicable to any material by providing an attenuation ratio for the vibration propagation based on the physics of vibration [19, 28].

While prior work has modeled vibration propagation for engineering applications, its relevance to VR haptic design has been largely overlooked. No existing system uses physics-inspired modeling of vibration propagation based on object material properties to modulate real-time haptic feedback. Scene2Hap introduces this missing link: it uses LLM-inferred material parameters to simulate spatially dependent attenuation, allowing vibrations to propagate through the virtual environment in a way that is perceptually grounded and responsive to user interaction.

3 Scene2Hap

Scene2Hap is an LLM-centered system that automatically designs object-level vibrotactile feedback for entire VR scenes, based on object semantics, physical properties, and spatial context. Its architecture is the first to use an LLM to extract information for haptic modeling from the VR scene, and uses this information for physics-inspired modeling for real-time user interaction. It operates at scale and without requiring manual authoring.

Scene2Hap begins by automatically extracting multimodal data from the existing VR scene. This data is used to drive a sequence of LLM components through prompt chaining. Their output provides the basis for two complementary strategies for haptic generation: (1) at startup, semantic descriptions of vibrating objects are used to retrieve or generate appropriate audios, which are then assigned to objects; and (2) during runtime, material properties, spatial relationships (represented as a contact graph), and pre-assigned audio signals are used to dynamically generate context-aware vibrations based on where the user is touching the scene. This architecture enables the generation of plausible vibration signals for each vibrating object in the scene, and it allows for vibration to propagate to neighboring objects in the scene based on spatial arrangement and material properties. When the user touches objects using standard VR controllers, the contextually correct vibration is generated in real-time. Overall, this transforms visual-only scenes into multimodal experiences that reflect both the physical and semantic structure of the environment.

In the following sections, we first describe *LLM-Based Haptic Inference*, which involves structured multimodal data extraction from VR scenes and prompt chaining using multimodal LLMs. This

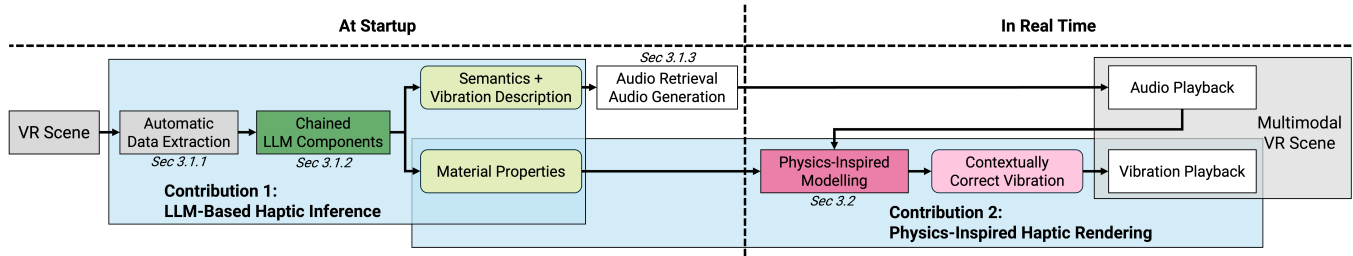


Figure 2: Architecture of Scene2Hap: (1) **At startup**, LLM-Based Haptic Inference uses a sequence of LLM components to automatically infer each object’s semantics, material properties, and vibratory behavior from multimodal information extracted from the VR scene. (2) **During user interaction**, Physics-Inspired Haptic Rendering generates contextually correct vibrotactile feedback in real-time based on the inferred objects’ vibratory behavior, material properties, and spatial configuration in the scene.

is followed by a brief description of how audio signals can be retrieved or generated. We then discuss how *Physics-Inspired Haptic Rendering* employs the information inferred in the previous step for physics-inspired generation of vibrotactile output based on user actions, estimated material properties, and the object’s context in the scene. A full system overview is shown in Figure 2.

3.1 LLM-Based Haptic Inference

Here we discuss the activities Scene2Hap performs at startup, once for the VR scene. This includes methods used for extracting data from VR scenes, and the tasks, purpose, and architecture of the chained LLM components we use, as each component is listed in Figure 2. The quality of the resulting data is evaluated in Section 5.1.

3.1.1 Automatic Data Extraction. We use multiple strategies to collect information at different levels of abstraction. This includes *Scene Information* (global context and aggregated scene attributes) to ensure that all LLM responses are contextually appropriate, and *Object Information* (properties of individual objects) used to generate material properties and inform the semantics of the object’s vibratory behavior. An overview of the data sources used can be found in Figure 3, in gray.

Scene Information. Scene2Hap collects two high-level inputs to characterize the overall environment: the *Scene Name* and a set of *Scene Images*. Since the *Scene Name* is developer-defined and often unreliable, we supplement it with *Scene Images*, which are screenshots captured from multiple angles within the VR environment. Details on these angles are provided in the Implementation section below.

Object Information. Scene2Hap collects visual and geometric information for each object to support semantic interpretation and material estimation. Two image types are used: the *Isolated Image*, showing only the target object from multiple angles, and the *Context Image*, captured from the same angles but including surrounding objects. In context images, a pink outline is automatically added to mark the target object to help the LLM visually disambiguate it. These visual inputs are used to identify what the object is and how it is used within the scene.

Alongside the images, Scene2Hap provides structured data. The *Object Name* is a predefined label provided by the developer of the VR scene, often ambiguous or generic. The *Size* consists of three numerical values in meters representing the dimensions of the object’s dominant surface—such as a tabletop of a table—by raycasting within the 3D mesh’s bounding box and calculating its median values. The *Relative Height* gives the vertical offset between the object’s bottom surface and the lowest object in the scene, which helps distinguish ambiguous elements—for example, identifying a flat surface labeled “Plane” as a floor rather than a wall or ceiling.

3.1.2 LLM Workflow and Components. While the data that can be automatically extracted from a VR scene is rich, it requires further processing to be useful for haptic feedback design. Humans can intuitively infer which objects might emit vibrations or what materials they are likely made of, but such information is not explicitly available in the raw data. To bridge this gap, we automate the final step of data enrichment needed for full-scene haptic authoring through prompting a multimodal LLM.

To inform the design of our architecture, we initially experimented with a straightforward, simple prompt (cf. Appendix A.1). This takes the above extracted information as input and directly prompts the LLM to estimate whether the object vibrates, to indicate its material properties (size, density, Young’s modulus, and Poisson’s ratio), and to describe the object’s vibration using a free-form sentence and keywords. We identified four major sources of incorrect results: 1) The LLM usually did not consider object semantics in the scene context. For instance, a miniature toy truck was judged to vibrate as a real truck, despite its small size and positioning on a desk. 2) Mechanical vibration originating from adjacent objects was wrongly considered as originating from the object itself. For instance, a mug on a desk was considered to vibrate, because “*Mugs can vibrate when placed on a vibrating surface or due to external forces*”. 3) Generated vibration descriptions were often ambiguous, especially missing “what object” and “how” it vibrates. 4) Object dimensions were often incorrectly assessed. Notably, zero thickness was often assumed for surfaces because Unity developers frequently use flat meshes with zero or close-to-zero thickness for boundary surfaces (e.g., floor, ceiling, wall). Informed by these findings, we developed the final prompting scheme to explicitly analyze object

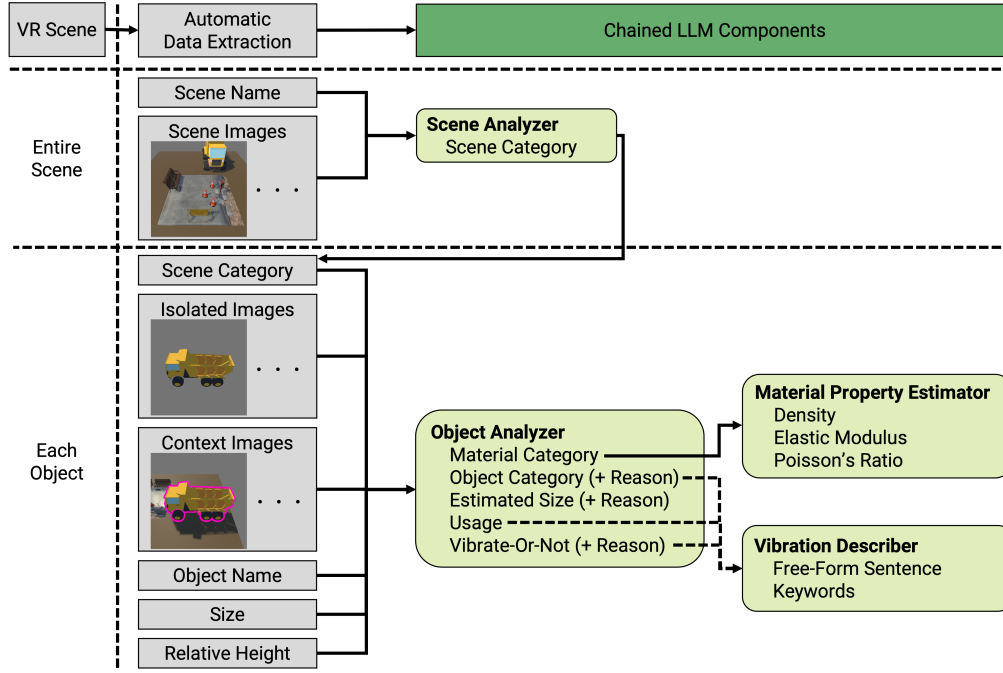


Figure 3: LLM-based haptic inference estimates the haptic properties of virtual objects by using an LLM workflow comprising four chained LLM components: Scene Analyzer, Object Analyzer, Material Property Estimator, and Vibration Descriptor. The images show one specific example.

semantics within the overall VR scene, to infer physically plausible dimensions despite potentially incorrect data in the VR scene, and to generate more specific vibration descriptions by explicitly considering the object category and its usage.

Figure 3 illustrates the final LLM workflow. To ensure simplicity and maintainability of the LLM-based system¹, our architecture adopts prompt chaining using four LLM components, each responsible for a specific subtask: the *Scene Analyzer* infers the global context of the scene, the *Object Analyzer* identifies relevant object semantics, the *Material Property Estimator* predicts likely material attributes, and the *Vibration Descriptor* generates corresponding haptic descriptors. Each component is prompted using a template (see Appendix A) that is automatically populated with multimodal information collected from the scene, as described in Section 3.1. The final output is a structured JSON object containing estimated semantic, material, and haptic properties, which is returned to the VR system for use during runtime haptic rendering. Next, we provide a detailed overview of all components:

Scene Analyzer. This component provides high-level contextual information for all subsequent prompts. It receives multimodal inputs: the *Scene Name* (a potentially ambiguous textual label) and multiple *Scene Images* (captured from different viewpoints in the VR environment). Based on these, it outputs a textual *Scene Category*, which represents the estimated type of scene (e.g., kitchen, office, workshop). This output is inserted into all downstream prompts to help improve their relevance and accuracy.

Object Analyzer. The purpose of this component is to infer detailed semantic and contextual information about each object, which is used in more refined downstream prompts. It receives a combination of structured and multimodal inputs. The *Scene Category*, as determined by the Scene Analyzer, as well as *Isolated Images* and *Context Images*. These are complemented by textual information including *Object Name*, *Size*, and *Relative Height*, as discussed in Section 3.1.1.

The Object Analyzer outputs several attributes: a textual *Object Category* and *Material Category*, an *Estimated Size* (in the same format as the input size) with adjustments if the raw values are implausible for the inferred category, a *Usage* description capturing how the object is likely used within the scene, and a boolean *Vibrate-Or-Not* label indicating whether the object should produce vibration. For selected attributes such as category, estimated size, and vibration status, the component also provides brief justifications. This additional reasoning is necessary for interpreting and validating the component’s outputs, especially in cases where corrections are applied. For example, it is common in VR scenes to model large boundary surfaces like walls or floors with near-zero thickness; in such cases, the LLM replaces physically unrealistic size values (e.g., 0.001 m) with plausible defaults consistent with the inferred object category.

The prompts for this component are designed to support reasoning based not only on object identity and scene context but also on physical plausibility, including the object’s realism – e.g., is it a toy car or a real car? – and its potential to vibrate. In particular, the component is allowed to infer vibration behavior even for inactive

¹<https://www.anthropic.com/engineering/building-effective-agents>

objects –such as power tools that are currently off – if it is likely that they would vibrate in interactive scenarios. This enables more complete coverage in haptic design. The output of this component is used to prompt the final two components and also provide the physics-inspired modeling step with information about object size.

Material Property Estimator. This component receives *Material Category* as input and outputs its material properties (*Density*, *Elastic Modulus*, *Poisson's Ratio*) in numerical values. These values are used to calculate vibration propagation in Section 3.2.

Vibration Describer. This component activates only for objects marked as vibratory (*Vibrate-Or-Not* = true). It receives the object's *Usage* as input and produces two types of textual outputs describing how the object should vibrate: *Free-Form Sentence*, used for audio generation, and *Keywords*, combining an *Object Category* and a verb, used for audio retrieval.

These outputs support the creation or selection of suitable vibration and sound profiles to match the inferred object behavior in context.

3.1.3 Audio Retrieval or Generation. The output of the *Vibration Describer* is used for identifying or generating an appropriate audio file for a vibrating object. We considered using dedicated text-to-haptics generation models (e.g., [77]); however, this would have increased the complexity of our system, as it would also require generating temporally synchronized audio, which is challenging. Basing the vibrotactile signals on the corresponding audio enables audio-tactile synchronization. We do not claim any contribution in this area, but instead build on the strong work by [45]. We describe the detailed process of audio retrieval/generation in the Implementation section below.

3.2 Physics-Inspired Haptic Rendering

So far, we have explained how Scene2Hap generates semantically and contextually appropriate descriptions of vibration, uses these descriptions to retrieve or synthesize matching audio files, and assigns these files to objects that were identified as vibration sources. However, this accounts only for the origin of vibration. In real environments, vibration propagates beyond its source and interacts with the surrounding materials.

To recreate this effect in VR, Scene2Hap uses a physics-inspired model for vibration propagation and attenuation that dynamically simulates how vibration travels through the scene. Vibration amplitude is highest near the source and attenuates with distance. Hard materials allow vibrations to travel farther, while soft materials dampen them more quickly. This behavior is essential for conveying information about material properties and spatial relationships between objects.

The exact propagation of vibration depends both on static properties—such as material type and object dimensions, inferred in the previous step—and on dynamic factors that must be computed in real-time. These include the position of objects, the user's point of contact with the scene, and the resulting attenuation of vibration amplitude along the propagation path. We now describe how Scene2Hap handles these real-time components, including runtime tracking of spatial configuration, and the application of the propagation model to compute localized haptic feedback.

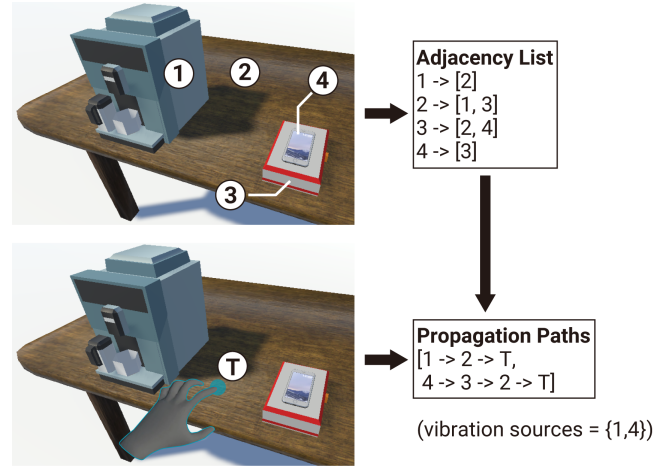


Figure 4: Physics-inspired haptic rendering builds the scene hierarchy and calculates all vibration propagation paths from vibration sources in real-time.

3.2.1 Real-Time Contact Graph. To identify spatial and physical relationships between objects in real-time, Scene2Hap builds and continuously updates a contact graph during runtime, treating the active VR scene as an undirected graph (Figure 4). In this graph, each virtual object is represented as a node, and an edge is created between two nodes when the corresponding objects are in physical contact.

When the user touches an object using a tracked VR controller (handled in Unity), Scene2Hap uses the current contact graph to identify all possible propagation paths from the touched object to known vibration sources. These paths are determined using a depth-first search through the graph. These paths are then passed to the vibration propagation model described in the following subsection, which calculates the vibration attenuation between neighboring materials based on material properties and spatial distance. This allows Scene2Hap to dynamically estimate the vibration amplitude at the point of contact, even as the user interacts with or moves objects during runtime.

3.2.2 Vibration Propagation and Attenuation. The vibration intensity at different points on a surface depends on how vibrations propagate through materials. While exact modeling of this behavior is complex, computationally intensive, and remains an active area of research in material science and physics, interactive VR systems require models that are efficient enough for real-time computation. To support haptic feedback at interactive frame rates, Scene2Hap uses a simplified yet physically grounded propagation model, building on the state-of-the-art [62]. As we will show in Study 2 below, this model offers a significant improvement over existing systems that do not include physics-based models.

According to [19, 28], the attenuation ratio can have an exponential behavior, depending on the material and geometrical properties of the surface. As we have estimates of these material properties from the LLM output, we can calculate this behavior. The attenuation ratio in point R with coordinates x and y can be calculated

as:

$$\Gamma(R) = e^{-k(R-R_0)} \quad (1)$$

where, R_0 is the location of applied vibration with coordinates x_0 and y_0 and k , the wavenumber, can be calculated using the following equation:

$$k^4 = \frac{\rho h \omega_0^2}{D} \quad (2)$$

Here, ρ and h are the density and the thickness of the surface, respectively and ω_0 is the angular frequency of the applied vibration at point R_0 . ω_0 is calculated by applying the Fast Fourier Transform to the audio file and finding its dominant frequency. D is the bending stiffness and can be calculated as:

$$D = \frac{Eh^3}{12(1-\nu^2)} \quad (3)$$

where, E and ν are the elastic modulus and the Poisson's ratio of the surface, respectively. All required parameters for these calculations are provided by the *Material Property Estimator*.

In Equation 1, k is a function of the surface's material and geometrical properties and $R - R_0$ represents the Euclidean distance between the touching point and the vibration source. Once the attenuation ratio is calculated for point R , it is used to scale the amplitude of the original vibration. This modulated amplitude is then used to drive the controller's vibrotactile feedback in real-time, matching the user's contact location in the VR scene. When multiple vibration sources are present in the scene, the sum of each attenuated signal is output as the final vibrotactile signal. This calculation is separately done for the position of each hand in the VR scene so that the user receives independent vibration feedback for each hand.

The specific model was chosen for its relative simplicity and suitability for real-time interaction. However, the Scene2Hap architecture can also support alternative models for simulating vibration propagation. For example, the scene could be modeled as a mass-spring-damper system to capture more complex dynamic behaviors.

Next, we explain the details of our specific implementation as used for our evaluation.

4 Implementation

We implemented the Scene2Hap concept in our system prototype as follows:

VR Experience. We implemented all VR scenes in Unity3D and ran them on a Meta Quest Virtual Reality headset. The Unity scene calculates the contact graph and propagation ratio in response to the user interaction at an interactive frame rate of 50 Hz, including the update of vibration signals. This calculation uses only Unity functions and is applicable to both static and moving objects.

Client/Server. Scene2Hap adopts a client-server model via HTTP communication. An HTTP client implemented in Unity collects multimodal information on VR scenes or objects and sends this information to the server built with the Python Flask framework. The server prompts the LLM to process this information to estimate semantics and physical context and sends its response back to the client. Our current implementation runs both the Unity client and Python server on a Windows 10 PC with an NVIDIA GeForce RTX

4090 GPU. All testing and evaluation for this paper were conducted on this machine.

Collecting Multimodal Information. We automatically collect multimodal information for the scene and each object within the Unity scene, as needed for haptic inference (Section 3.1). Textual (*Scene Name*, *Object Name*) and numerical (*Size*, *Relative Height*) information is obtained by accessing each object's metadata. For *Isolated Images*, the system automatically moves the camera object in the Unity scene and takes eight images of the entire target object at a 45-degree angle from above and below, rotating the camera in the horizontal plane by 90-degree increments. The camera renders only the target object. For *Context Images*, the system takes the target object with surrounding objects at the same angles as *Isolated Images*. Here, the system casts a ray from the camera object to the target object and culls objects that hit the ray. For *Scene Images*, the system takes four images at a 45-degree angle from above, rotating the camera in the horizontal plane by 90-degree increments. Here, the system calculates the center of all the existing objects in the scene. Similar to *Context Images*, the system culls objects that are between the camera and this center point. This culling process sometimes results in unnatural culling of nearby objects or, conversely, in showing objects in the way, due to the variety of the scene arrangement. While it is not critical in our evaluations, this issue could be addressed in future work with a more advanced algorithm. We used a Unity asset² to add a pink outline to the object in context images. The outline sometimes does not completely enclose the object, depending on the setting of the target object in a VR scene. To deal with this case, the LLM is instructed to focus on the object in context images that most resemble the object shown in isolated images.

LLM Workflow. We implemented an LLM workflow inside the Python server using the OpenAI API (GPT-4o, the model temperature was set to 0.2) and the LangChain framework³. The multimodal information sent from the Unity client is organized into a format that can be fed into this LLM workflow. LLM-based haptic inference takes around 9–12 seconds per object in the current setup, based on Study 1's measurement result.

Audio Retrieval/Generation. We used Freesound API⁴ for an external web-based audio database in audio retrieval and used AudioGen [45] (model = AudioGen-Medium-1.5B⁵) for a text-to-audio model in audio generation, running on the same PC as the Unity client and Python server do. AudioGen requires a GPU with at least 16 GB of memory⁶.

In this work, Scene2Hap first tries to retrieve up to 5 best-matching audio files by querying the Freesound database with the *Keywords* generated by the *Vibration Describer* LLM component. If no file matches these keywords, Scene2Hap generates 5 audio files by feeding the *Free-Form Sentence*, generated by the same component, to the AudioGen model. The system automatically removes silent sections at the beginning and end of the audio and keeps only samples longer than 2.5 seconds. Out of these candidates, the

²<https://assetstore.unity.com/packages/tools/particles-effects/quick-outline-115488>

³<https://www.langchain.com/>

⁴<https://freesound.org/>

⁵<https://huggingface.co/facebook/audiogen-medium>

⁶<https://github.com/facebookresearch/audiocraft/blob/main/docs/AUDIOGEN.md>

finalist is selected as follows: To favor continuous sounds, the system discards audio files with a dynamic range of more than 5 dB, as these would be less well-suited for continuous looping (if no candidate is remaining, this step is iteratively repeated, increasing the threshold by 5 dB until a solution is found). Finally, the system selects the most harmonic audio, i.e., the audio file with the lowest spectral flatness, because high harmonic content was typically more pleasing than noisy files. The amplitude of this audio file is normalized.

We believe that there are many promising approaches in development and that this step will become trivial in the near future.

Audio Processing. For simplicity, Max/MSP was used for audio processing. Audio and vibration playback times are synchronized between Unity and Max/MSP by sending a UDP message when each audio file is played in the Unity scene for the first time. Both the audio and vibration are played in a loop. Once the material and geometrical properties of the scene objects are identified by the LLM, the attenuation ratio is calculated based on the coordinates of the point the user touches in Unity. Every time this attenuation ratio of a vibration source is updated, Unity sends a UDP message to Max/MSP, including the attenuation ratio, the path of the audio file, and a hand index (i.e., 0 for left hand, 1 for right hand). The audio signal is converted to a vibrotactile signal by limiting the frequency spectrum of the applied vibration to the human tactile sensitivity band. Since Pacinian corpuscles are mostly responsible for acquiring vibrations on the skin, we applied a state variable filter (configured as a band-pass filter) with a resonance frequency at 250 Hz, corresponding to the peak detection frequency of the Pacinian corpuscles [21]. This filtering reduces the complexity of the calculation with minimal impact on perceived quality (see Study 2). Therefore, the low and high-frequency components of the applied vibration signal were filtered out. In addition, the vibration amplitude was modulated using a simplified approach given in Section 3.2. For simplicity, in this work, we calculated this propagation when the hand directly contacts a vibration source (attenuation ratio is 1) or there is only one intermediate object between the hand and a vibration source (e.g., feel the vibration of a phone from the desk where it is placed). The final vibration is calculated by summing all the attenuated vibration signals in Max/MSP. While this simple approach was effective for our system, we acknowledge that more advanced conversion methods exist (e.g., perception-level translation [46], frequency shifting [60], or pitch matching [41]), which remain an interesting avenue for future exploration.

Haptic Device. The system renders vibrotactile signals using two Tachammer Drake HF vibrotactile actuators, one attached to the handle of each Meta Quest VR controller using tape and fixed with zip ties, such that the user can feel vibration on both hands while interacting in the VR scene. The vibrotactile signals were generated in Max/MSP and amplified using a Visaton 2.2LN Amplifier.

5 Evaluation

To validate Scene2Hap, we conducted three studies investigating (1) the capability of LLM-based haptic inference, (2) the effect of physics-inspired haptic rendering on the user's haptic perception, and (3) the overall experience in a full VR scene.

5.1 Study 1: LLM-Based Haptic Inference

This study aims to evaluate how correctly the proposed LLM-based haptic inference can infer the attributes of each object in the scene. We evaluate this in two ways: (1) attributes that leave room for subjective interpretation (scene, object, and vibration description) were evaluated by human raters in an online study; (2) attributes that could be assessed objectively (physical material properties including density, elastic modulus, Poisson's ratio) were assessed by comparing to known data from the literature.

VR Scenes. For this study, we have selected six Unity scenes downloaded from the Unity Asset Store⁷, as shown in Figure 5. We selected these scenes following three main criteria: (1) the scene includes multiple objects that are likely to vibrate and others that are not likely to vibrate; (2) the scene depicts a realistic setting, including typical objects that are commonly used in this setting; (3) the scene includes a moderate number of virtual objects (< 50) to remain feasible within the scope of the study. The selected scenes cover diverse VR scene settings and corresponding objects: bathroom, kitchen, construction site, and garden. We kept the scenes unmodified, but made three minor changes: First, we adjusted the overall scale of each scene to be close to that of real-world environments, as usually done by VR designers, helping our system correctly judge the size information of each object. Second, we added a small number of objects to evaluate the capability of our LLM-based haptic inference in understanding object semantics with even more challenging conditions (i.e., Pan and Truck in different usage contexts, respectively) or to make the scene appear more natural (i.e., brown plane in Construction Site scene), as summarized in Figure 5. Third, we removed a few common objects that were already present in another scene and combined some unconventionally divided elements (e.g., individual floor panels) into a single object (floor) to focus on a more meaningful evaluation.

We used two scenes (Bathroom 1, Kitchen 1) as scenes during development, for iteratively developing our LLM architecture and refining the prompts. The remaining four scenes were held out and remained unseen to be used for the evaluation of the system. None of the objects in the development scenes were included in the test scenes. We also measured the time to complete the whole haptic inference in each scene, as shown in Figure 5.

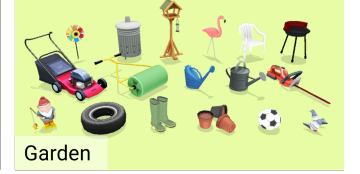
Online Questionnaire. We created an online questionnaire that gathered participants' subjective assessment of how correct they considered the results of LLM-based haptic inference on the test scenes to be. Participants assessed both the correctness of results pertaining to the entire VR scene and to the individual objects in the scene. For each scene, the questionnaire provided the multimodal information that was actually fed into the LLM (*Scene Name*, four *Scene Images*). It instructed the participants to rate the correctness of the estimated *Scene Category* on a 5-point Likert Scale (1=fully incorrect – 5=fully correct). Similarly, for each object, the questionnaire provided multimodal information fed into the LLM (*Scene Category*, eight *Isolated Images*, eight *Context Images*, *Object Name*). Participants had to rate on a 5-point Likert Scale the correctness of the estimated *Object Category + Reason*, *Material Category*, *Usage*, and *Vibrate-OR-Not + Reason*. If *Vibrate-OR-Not* was true, we

⁷<https://assetstore.unity.com/>

Development Scenes



Test Scenes



	Scene	Scene Name (defined by the downloaded scene)	Number of Objects	Modifications	Estimated Scene Category	Correctness Rating (Avg (SD) on 5-point Likert scale)	Processing Time (in s)
Development	Bathroom 1 [74]	bathroom_ and_laundry_ showcase	34	—	Bathroom	—	—
	Kitchen 1 [35]	Demo	20	Added a <i>pan</i> <i>roasting meat</i> on the heater, a <i>pan</i> and a <i>miniature refrigerator</i> on the desk. Removed a <i>washing machine</i> and a <i>dryer</i> that were already included in this scene.	Kitchen	—	—
Test	Bathroom 2 [1]	Demo	36	Added an <i>electric toothbrush</i> and a <i>hair dryer</i> . Combined separate ceiling, wall, and floor <i>panels</i> into one object.	Bathroom	4.90 (0.32)	415
	Kitchen 2 [75]	Presentation	44	Added a <i>pan</i> and a <i>toy truck</i> on the desk. Combined separate floor <i>panels</i> .	Kitchen	5.0 (0.00)	491
	Construction Site [54]	DemoScene	14	Added a <i>truck</i> and a <i>brown plane</i> under the truck as a ground.	Construction Site	4.70 (0.48)	132
	Garden [59]	objects1	23	—	Backyard	4.10 (1.10)	237

Figure 5: Two scenes were used for developing the prompts for our LLM-based haptic inference module, and four test scenes for evaluation. The table includes detailed information for each scene. Empirical results show that the scene category was correctly estimated regardless of the inappropriate scene names defined in the downloaded scenes. Processing time is the time required to complete the LLM-based haptic inference for the entire scene.

	Object Category (Avg (SD))	Material Category (Avg (SD))	Usage (Avg (SD))	Vibrate-Or- Not (Avg (SD))	Free-Form Vibration Description (Avg (SD))	Keyword Vibration Description (Avg (SD))
All Objects (30)	4.51 (1.21)	4.21 (1.26)	4.51 (1.20)	4.12 (1.42)	3.63 (1.47)	3.61 (1.52)
Correct Objects (24)	4.88 (0.48)	4.33 (1.13)	4.83 (0.59)	4.61 (0.87)	4.30 (0.86)	4.34 (0.90)
Hard-to-Judge Objects (3)	4.80 (0.48)	4.80 (0.48)	4.83 (0.38)	3.23 (1.48)	3.07 (1.46)	2.90 (1.45)
Incorrect Objects (3)	1.27 (0.91)	2.63 (1.61)	1.57 (1.38)	1.10 (0.40)	1.30 (0.79)	1.13 (0.35)

Table 1: Participants’ ratings of the correctness of LLM output on a 5-point Likert scale (1=fully incorrect – 5=fully correct). The results show that the LLM-based haptic inference successfully infers the semantics of diverse virtual objects in alignment with human raters for most objects.

further asked participants to rate the correctness of the *Free-Form Sentence* and of the *Keywords* for vibration description. As the total number of objects present in all scenes would have exceeded the scope of the questionnaire, we selected a total of 30 objects that met

at least one of these criteria: (1) the object was estimated to vibrate in the scene, (2) the object is in contact with another object that is estimated to vibrate, or (3) the object was added or modified by the experimenter for a deeper evaluation of our proposed architecture,

as summarized in Figure 5. The questionnaire did not cover *Size*, *Relative Height*, or *Estimated Size* because these numerical values were hard for the raters to judge intuitively from images alone.

Participants. We recruited 10 participants (aged 24 to 34; 6 identified as male, 4 as female). The online rating procedure took approximately one hour.

Results and Discussion. Participants rated the LLM-inferred *Scene Categories* as highly correct (AVG = 4.68, SD = 0.67). Considering that most of the *Scene Names* defined by the downloaded Unity scenes did not match the estimated scene category, as seen in Table 1, this result indicates that the *Scene Analyzer* component can effectively estimate the semantics of VR scenes based on their multimodal information.

Table 1 shows an aggregated view of the correctness ratings of the LLM-inferred object properties. In Supplemental Materials, we share a detailed list, comprising the LLM-inferred properties and the corresponding participant ratings for all individual objects. The first row of Table 1 lists the average response for all objects. Averaged over all objects, the ratings indicate that the LLM could correctly infer *Object Category*, *Material Category*, *Usage*, and *Vibrate-Or-Not*, with scores well beyond 4 on the 5-point Likert scale. The free-form and keyword descriptions of vibrations received slightly lower ratings of around 3.6.

A deeper analysis of individual objects revealed that objects can be subdivided into three main clusters: for the vast majority of objects (24/30, see Supplemental Materials), the LLM has provided *correct* results that on average were rated between correct (4) and highly correct (5) for all items (average ratings between 4.30–4.88). This includes the demanding semantic assessment of whether the object can vibrate or not (AVG = 4.61, SD = 0.87) and the description of the vibration using free text (AVG = 4.30, SD = 0.86) and keywords (AVG = 4.34, SD = 0.90). These results indicate the *Object Analyzer* and *Vibration Describer* components can effectively estimate the semantics of various objects based on their multimodal information, including new types of scenes (Construction Site, Garden) that were not used during the development process.

A second, much smaller cluster comprised 3 objects that were clearly *incorrectly* estimated by the LLM (average ratings between 1.10 – 2.63), see the last row of Table 1. All these 3 objects were variations of the same Hand Towel Rack object present in the Bathroom 2 scene with the same geometry. Their object categories were mistakenly estimated as "shower head" or "hair dryer" probably due to their similar appearance when looking at the object from the side. This misrecognition of the object category has led to consistently low ratings by all participants for all items belonging to these 3 objects.

A final cluster comprised three objects that turned out to have vibration properties that are *hard-to-judge* even for humans. Participants were split in their assessment of whether these objects ("frying Pan" and "saucepan" objects on the heater and "stove oven" in Kitchen 2 scene) should vibrate or not in the scene. This led to very differing ratings with a high standard deviation for *Vibrate-Or-Not* (AVG = 3.23, SD = 1.48) and for both vibration descriptions of *Free-Form Sentence* (AVG = 3.07, SD = 1.46) and *Keywords* (AVG = 2.90, SD = 1.45). This split was somewhat expected because these

cases were hard to judge: whether pans vibrate on the heater depends on multiple factors, such as the heat intensity, what food or liquid is to be cooked, and in what quantity. Note that for the other properties (*Object Category*, *Material Category*, *Usage*), the *Object Analyzer* component worked appropriately even in the case of *hard-to-judge* objects.

Interestingly, the LLM was able to understand the advanced semantics of virtual objects in scenes. For instance, the truck object with the same (but differently scaled) 3D model was successfully recognized as a "dump truck" in the Construction Site scene and as a "miniature toy truck" in the Kitchen 2 scene, probably because our module explicitly made the LLM consider object semantics based on its size. Also, the pan object with the same geometry in the Kitchen 2 scene was recognized as "it should vibrate" when placed on the heater because "*The frying pan may vibrate slightly due to the heat from the stove when cooking*". In contrast, it was recognized as "it should not vibrate" when put on the table because "*A frying pan does not typically vibrate unless it is on a heat source*". The authors consider this LLM's response a correct assessment, and considering the split ratings provided by study participants, this may indicate the LLM can even outperform human judgment in selected cases. Overall, these findings show the great potential of LLMs in automatic haptic design that considers the diverse semantics of objects.

Correctness of LLM-estimated Material Properties. To highlight that the data of material properties provided by GPT-4o is reasonably good (as also indicated from the literature [89]), we compare its output with known data from the literature. For this analysis, we first collected a pool of materials from Ref. [19]. We then asked GPT-4o to "Choose ten of them that are primarily used in everyday life applications". Then, we asked it to "give the Density (kg/m^3), Elastic Modulus (N/m^2), and Poisson's Ratio of the selected materials".

Table 2 shows the results. To reflect real-world variability in material properties and corresponding variability of measures, data from Ref. [19] is supplemented with data from other sources. The results show that the LLM's estimations are overall within the range of measures from the literature. In cases where this is not the case, deviations are minor – the deviations identified in our sample are highlighted in red. Overall, this suggests that the values provided by GPT-4o are a reasonable foundation for basic physical modeling.

5.2 Study 2: Effect of Vibration Propagation on Haptic Perception

To understand the effect of vibration propagation and attenuation, we performed a second evaluation study demonstrating that the attenuated vibration improves user experience, may help in perceiving materiality, and clearly supports users in their spatial awareness of the scene. While Scene2Hap uses both LLM prompting and physical modeling, the goal is not exact replication or naturalistic modeling, but to enhance user experience in virtual environments. In addition to generally improving user experience, we hypothesize that the correct propagation of vibration improves the perception of materiality and supports spatial understanding. To test this, we evaluate the effect of attenuated vibration propagation on user experience in three scenes.

Material	Literature			GPT-4o		
	Density (kg/m ³)	Elastic Modulus (N/m ²)	Poisson's Ratio	Density (kg/m ³)	Elastic Modulus (N/m ²)	Poisson's Ratio
Aluminum	2570 – 2950 [52]	68 – 88.5 × 10 ⁹ [52]	0.32 – 0.36 [52]	2700	69 × 10 ⁹	0.33
Steel	7820 – 7860 [8]	190 – 210 × 10 ⁹ [8]	0.27 – 0.32 [8]	7850	200 × 10 ⁹	0.30
Copper	8930 – 8940 [51]	121 – 133 × 10 ⁹ [51]	0.34–0.35 [51]	8960	110 × 10 ⁹	0.34
Glass	2400 – 2600 [68]	50 – 130 × 10 ⁹ [68]	0.15 – 0.3 [6, 27]	2500	70 × 10 ⁹	0.23
Plywood	400 – 600 [81]	7 – 8.6 × 10 ⁹ [81]	0.2 – 0.3 [24]	600	10 × 10 ⁹	0.3
Gypsum Board	545 – 700 [18]	0.47 – 2.5 × 10 ⁹ [18]	0.24 [53]	850	2.5 × 10 ⁹	0.25
Brick	1900 – 2200 [19]	6 – 14 × 10 ⁹ [58]	0.1 – 0.25 [44]	1920	12 × 10 ⁹	0.20
Asphalt	1800 – 2300 [19]	0.5 – 140 × 10 ⁹ [7]	0.3 – 0.4 [11]	2300	1 × 10 ⁹	0.35
Oak	700 – 1000 [19]	2 – 10 × 10 ⁹ [19]	0.33 [71]	700	12 × 10 ⁹	0.30
Plexiglass	1120 – 1150 [19, 87]	4.5 – 5.6 × 10 ⁹ [19, 34]	0.35 [34]	1180	3.3 × 10 ⁹	0.35

Table 2: Comparison of LLM’s material property estimation to established measures from the literature. The values provided by GPT-4o fall within the ranges identified from the literature. In cases where they do not, the value is indicated in red.

We assess haptic experience in three ways: (1) general usability, using the haptic experience design framework by Kim and Schneider [43] (utility, causality, consistency, saliency); (2) perceived materiality; and (3) spatial awareness. Each is addressed through targeted questionnaires following interaction with the system.

Participants. We recruited 10 participants (aged 20 to 34; 6 identified as male, 4 as female; 9 right-handed, 1 left-handed). Each study session took approximately one hour.

Scenes. Each participant experienced three VR scenes, each with a respective audio and vibration source: a speaker playing loud music on a metal table, a smartphone buzzing on a wooden table, and a washing machine running on a wooden floor (see Figure 6). This study used simplified scene settings to focus the evaluation on the vibration propagation and attenuation. The study used the same audio files and material parameters pre-generated by Scene2Hap for all participants.

To ensure controlled evaluation, we defined three touch points on the tables (for the first two scenes) and the floor (for the third) at distances of 0.0 (vibration source), 0.4, 0.8, and 1.2 m from each source. Participants were free to reach and touch any of the designated points. Each scene played its respective audio continuously with constant intensity, and the synchronized vibration was played when the participant touched a point. Scenes were presented in blocks, with randomized order.

Tasks and Questionnaire. Participants were instructed to touch each point under three conditions: *no propagation* (only the vibration source vibrates), *full propagation* (surfaces vibrate with the same amplitude as the source), and *attenuated propagation* (surfaces vibrate with attenuated intensity based on material and location). Each participant experienced all three conditions within each of the three scenes, resulting in nine trials in total. The order of scenes and the order of conditions within each scene were randomized.

After each condition, participants completed a questionnaire using Semantic Differentials [4], with immediate ratings collected after each trial (here, a high score indicates that participants agree, and a low score indicates that they disagree; please refer to Appendix B for the exact questions). They responded to six items using a continuous bipolar scale. The first four items – utility, causality, consistency, and saliency – were drawn from Kim and Schneider’s

framework [43] and served to assess general usability. The final two items – materiality and spatial awareness – were added to address our specific hypotheses. The exact questions can be found in the appendix.

Procedure. First, participants provided informed consent and demographic information. They then completed a familiarization session, where they freely explored all three scenes and all three haptic conditions. The trial session lasted as long as needed for each participant to feel comfortable with the setup. In the main experiment, participants performed free exploration in each scene under each of the three haptic conditions. The participants put on a noise-canceling headphone and listened to the audio played back in each VR scene. After each condition, they completed a questionnaire. Participants were allowed to take breaks between scenes if desired. The study took place in our lab and lasted approximately 45 minutes. Participants received compensation at the rate typical for our institution.

Results and Discussion. Average results and confidence intervals for this study are shown in Figure 7. We will next provide the results for Usability, Materiality, and Spatial Understanding.

Usability

Looking at the descriptive statistics in Figure 7 highlights that the *attenuated propagation* condition was rated highest for all Usability Requirements, and *no propagation* was rated lowest.

The data showed substantial deviations from normality in most cells (Shapiro–Wilk between $W = .69$ – $.95$, many $p < .01$), homogeneity of variance (Levene’s test, $F(11, 348) = 2.16$, $p = .016$). Visual inspection suggested that these violations were mainly due to responses clustering at the endpoints of the semantic differential scales, resulting in skewed distributions.

We therefore used a repeated measures ART-ANOVA [88] to examine the effects of Propagation (No Propagation, Full Propagation, Attenuated Propagation) and Usability Requirement Measure (utility, causality, consistency, saliency) on participant ratings, averaged across scenes. Bonferroni-corrected comparisons showed that Attenuated Propagation was rated significantly higher than No Propagation ($p < .001$) and Full Propagation ($p < .001$). There was also a significant difference between No Propagation and Full Propagation ($p = .022$). This indicates that, overall, the conditions

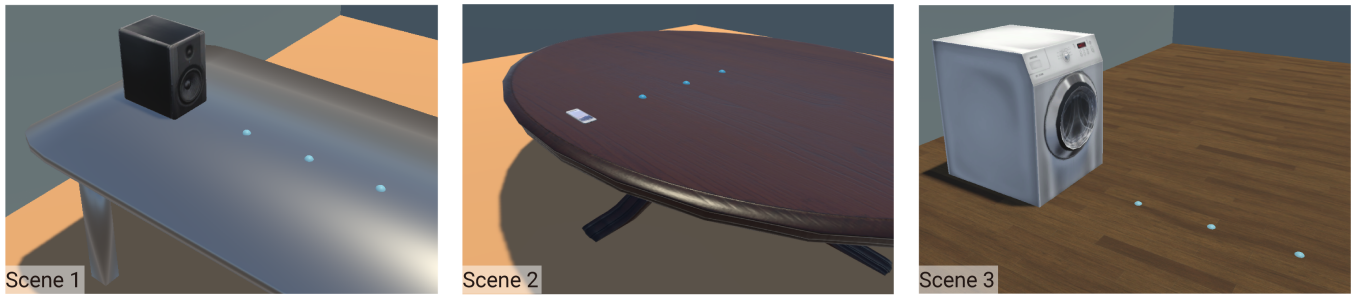


Figure 6: Study 2 used three scenes to investigate the effect of our physics-inspired haptic rendering on VR haptic perception.

with *attenuated propagation* lead to the highest usability, and that this effect was measurable for all Usability Requirements.

Materiality

Descriptive statistics show that all ratings for *Materiality* were comparatively low, however the condition with *attenuated propagation* was clearly rated strongest (see Figure 7).

As the data did not meet the normality criterion (Shapiro–Wilk, $W = .8228$, $p = .0274$ for attenuated propagation), the non-parametric Friedman test was used to examine the effect of *Propagation condition* on participant responses, averaging over the scenes.

The analysis revealed a significant main effect of *Propagation condition*, $\chi^2(2) = 8.22$, $p = .016$ with a moderate effect size (Kendall’s $W = .41$), indicating that participant responses differed significantly across the different propagation types. Bonferroni-corrected Wilcoxon test identified a significant difference between full propagation and attenuated propagation ($p = .0117$).

The overall low materiality scores may be attributed to general limitations of current VR systems (e.g., no force feedback when touching an object). However, even though the overall ratings were relatively low, the propagation of vibration through the scene significantly improved participants’ experience of materiality.

Spatial Awareness

Finally, we found strong differences in how propagation conditions affected the ratings of *Spatial Awareness*. Users rated the *attenuated propagation* highly, and the *no propagation* condition low (see Figure 7).

As, again, the data did not meet the normality criterion (Shapiro–Wilk, $W = .7932$, $p = .0120$ for attenuated propagation), the non-parametric Friedman test was used to examine the effect of propagation conditions on *Spatial Awareness*. The analysis revealed a significant main effect of *Propagation condition* ($\chi^2(2) = 12.80$, $p = .002$), with a large effect size (Kendall’s $W = .64$).

Post-hoc Wilcoxon tests with Bonferroni correction showed that Attenuated Propagation received significantly higher ratings than both Full Propagation ($W = 1$, $p = .012$, $r = .85$) and No Propagation ($W = 1$, $p = .012$, $r = .85$). The comparison between Full Propagation and No Propagation was not significant ($W = 6$, $p = .082$, $r = .69$).

This highlights that participants experienced a strong improvement in spatial awareness through the attenuated vibration.

In summary, our findings show that attenuated vibration propagation improves user experience. Usability Requirements (utility,

causality, consistency, saliency) showed consistent improvements, and perceived materiality was rated significantly higher under attenuated propagation. The strongest effect was observed for spatial awareness, with attenuated propagation clearly enhancing participants’ reported understanding of the spatial arrangement of the VR scene. These results demonstrate that physics-inspired haptic rendering supports both material and spatial perception in virtual environments.

In exit interviews, participants highlighted the immersive benefits of spatially dynamic vibration. For example, P3 shared that “*The condition with the propagation was realistic... it was pretty much real-world, I really liked it, and that was my favorite.*” Participants also immediately saw useful applications in VR, for example, in searching for the source of a vibration. However, participants noted that the propagation of vibration might not be as relevant if there are strong multimodal cues, e.g., P2 noted that “*The scene with the speaker was not really noticeable because the sound was so loud, and haptics... I could not differentiate.*”

5.3 Study 3: User Experience in Full VR Scene

To investigate the overall user experience in a full VR scene, we collected feedback from participants who interacted with the VR scene designed with our end-to-end pipeline.

Participants. We recruited 10 participants (aged 23 to 34; 5 identified as male, 5 as female). Each study session took approximately one hour.

Scene. We used a modified version of the Kitchen 2 scene from Study 1. To evaluate our pipeline in an interactive and realistic VR experience, we modified the scene in terms of the following three points: (1) We added four objects (a handheld fan⁸, an electric toothbrush⁹, a smartphone¹⁰, and a washing machine from the Bathroom 1 scene in Study 1). The first three objects are portable, allowing the user to grab or release them with the VR controllers and move them in the scene. (2) We removed three objects (two pans on the oven and a fried egg on a pan) due to the lack of consensus in Study 1 on whether they should vibrate in the scene. (3) We made the scene look more natural by adding walls and a window¹¹, and adjusting the floor size. We ran LLM-based haptic inference once

⁸<https://skfb.ly/oStXz>

⁹<https://skfb.ly/pwqpM>

¹⁰<https://assetstore.unity.com/packages/3d/props/electronics/free-smartphone-90324>

¹¹<https://assetstore.unity.com/packages/3d/environments/apartment-kit-124055>

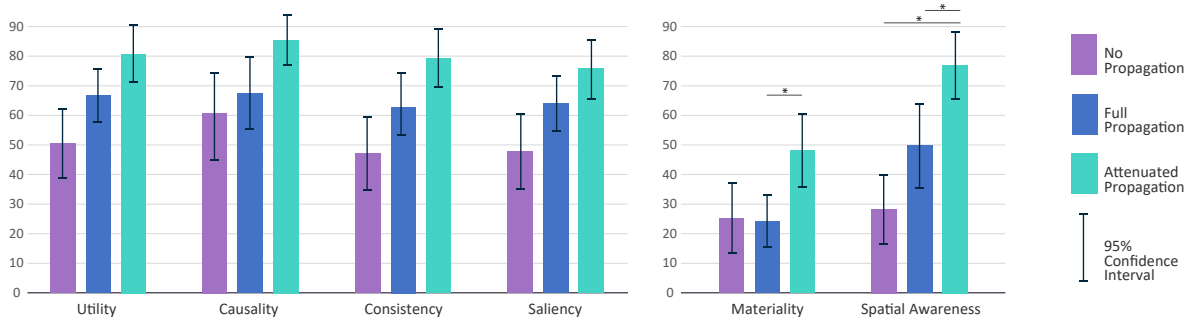


Figure 7: Results of Study 2. Purple, blue, and cyan colors show the average ratings of the participants for no, full, and attenuated propagation types, respectively. In this experiment, four usability requirements (utility, causality, consistency, saliency) as well as perceived materiality and spatial awareness were probed. Significant differences from the post hoc analysis are indicated by * if $p < .05$ and by ** if $p < .001$.



Figure 8: Study 3 used a modified version of Kitchen 2 scene (left). The vibration sources were set to the seven objects (right).

for this scene in advance and used the common material properties and audio signals for all participants to ensure comparability. The vibration sources were set to seven objects: handheld fan, electric toothbrush, smartphone, washing machine, refrigerator, cooker hood, and microwave.

Tasks and Questionnaires. To ensure all participants explored the relevant aspects of the scene, we asked them to follow a shared procedure. (1) They observed the scene visually for 30 seconds. (2) They were asked to identify which devices were turned on - they were encouraged to walk, touch, and continue exploring until all seven active devices were found. (3) They were asked to explore how lifting, moving, and combining portable vibrating objects affected the resulting vibration: they examined table and counter surfaces with and without mobile objects, tested vibration transfer by placing multiple devices on surfaces, and compared sensations of devices held in hand and placed on objects.

After completing the tasks, participants filled out two questionnaires using 5-point Likert scales. The first (from Study 2) assessed

usability (utility, causality, consistency, saliency), materiality, and spatial awareness. The second addressed realism, immersion, presence, clarity of feedback, engagement, and satisfaction (Appendix C). Likert items were chosen instead of semantic differential scales, as we do not intend to make statistical comparisons between conditions. An exit interview was conducted with each participant for the qualitative analysis. These interviews were structured by the questionnaires: the participant and experimenter together iterated through all the items and together reflected on why the participant chose the respective response to it.

Procedure. Participants signed a consent form, gave demographic information, and were briefed that some devices in the kitchen would produce audio and vibration. They explored freely while following task instructions. Afterward, they completed both questionnaires and took part in a recorded qualitative interview discussing each questionnaire item.

Data and Analysis. A chi-square test of independence indicated a significant association between questionnaire items (13) and response types (5), $\chi^2(48) = 83.20$, $p = 0.0012$, suggesting systematic response patterns. Qualitative data were transcribed and cleaned of non-verbal utterances. Statements by participants were extracted. The rest of the analysis was conducted using affinity clustering. Initially, one author assigned clusters per question, which were used to better understand the Likert items, and then assigned clusters across questions. In a collaborative coding session, these clusters were reviewed together by three authors and edge cases were discussed. One author completed the affinity clustering.

Qualitative Results Discussion. Figure 9 shows the participants' questionnaire responses. Overall, participants agreed on the positive contribution of haptic feedback. The ratings show that Scene2Hap positively contributes to users' perception of spatial awareness, realism, feedback clarity, causality, and satisfaction; those dimensions received only positive ratings. Engagement, saliency, presence, and utility were rated positively by a clear majority of participants. In contrast, materiality received mixed responses. The interviews provided insights on the ratings: Some participants felt that vibration added to their sense of materiality – for example, understanding the material of a table when vibrating objects were placed on it. In contrast, others found that the added vibration highlighted the lack of materiality in other aspects. For instance, the absence of normal forces and the inability to distinguish between surfaces such as plastic or wood became more apparent once haptics was added for vibration sources.

5.3.1 Resulting Themes. One cluster that emerged was around **comparisons to the real world**. Many participants emphasized that vibrations felt authentic in comparison to similar real-world sensations, such as a phone vibrating. Presence was stronger when vibrations carried through solid objects, as in real life. P6 stated: *“When I hear a vibration, like my phone alarm ringing, I feel it when I reach out, and that is how life is supposed to work. The same with the toothbrush and the fridge; in my head, things should work like this, and that is what realism is.”* Still, some participants (P1, P9, and P6, who otherwise appreciated the realism) highlighted that vibrations sometimes felt too similar and less varied than expected from real life. A strong anecdote came from P1, who experienced an incident where they almost fell because they tried to support their weight on a virtual chair. Reflecting on this experience, they suggested that vibration may have subconsciously made them believe the virtual world was more real: *“I dropped the toothbrush, I tried to use the chair to get up, but there was no chair. [...] being in the virtual environment felt real.”*

The **propagation of vibration** was a major discussion point. Generally, participants—without being explicitly prompted—noticed and praised the propagation mechanism. It was described as making the experience feel more realistic and increasing participants' sense of presence in the scene. Participants stated that as they moved their hands and approached objects, they could identify the vibration source. For example, P8 said: *“With three items on the table, as I approached them the vibration got stronger and clearer, helping me identify the source.”* Propagation also improved spatial awareness by helping participants understand their distance from the source. As P7 described: *“The table had vibration. As I moved towards the object,*

vibration got stronger, and as I went away, it got weaker. That gave me spatial awareness.” Participants mentioned that propagation attracted their attention to vibrating devices, unlike non-vibrating ones, such as the pan, which they paid less attention to.

Many participants used **comparisons between objects** to discuss haptic feedback. For example, participants 1, 4, 5, and 9 highlighted that vibration helped them distinguish between objects. However, participants 2, 3, and 6 felt that this aspect might be improved, stating that larger and smaller objects felt too similar. Interestingly, P9 reported that in one situation they could not differentiate between the fan and the washing machine—here the fan was placed on the washing machine, and the vibration the user felt was the vibration of the washing machine propagating through the fan. Participants also highlighted that while they could distinguish the behavior of devices tactually, the vibration *“didn't reveal whether it was metal, wood, or rubber”* (P4).

Another cluster that emerged was around discussions of haptics in the context of **other senses**. Many participants highlighted the importance of multimodality while praising the haptics. For example, P5 stated: *“Vision, hearing, and touch all need to work together. The most important haptic feedback was touching the table and sensing feedback from devices. Vision alone could not show differences, but haptics let me understand which devices were on and distinguish between objects like the toothbrush and cell phone.”*

A final topic that came up was the role of **personalization** of vibration. For example, P1 would have preferred that they could turn off the phone's vibration. Participants 8, 9, and 10 all expressed the wish for personalization options.

In summary, we see this as a strong endorsement for our Scene2Hap approach. Participants praised the immersion that the haptics provided. Not only was the vibration generally highlighted as realistic and useful, but participants also successfully manipulated vibrating objects and could feel how re-positioning and combining objects affected how vibration propagated through the scene in real time. Critical points were raised regarding some vibration signals that were felt too similar—this suggests that the fixed cutoff frequency chosen may not have been ideal. Similarly, the richness in vibration from active objects also made participants more critically aware of the missing haptic feedback from passive objects. The wish for personalization also highlights that additional editing or regeneration abilities would be a valuable addition in future iterations.

6 Discussion

The three evaluations jointly demonstrate the effectiveness and robustness of Scene2Hap. Study 1 confirmed that the system can accurately infer semantic and physical attributes of virtual objects based on automatically extracted multimodal scene data. This includes nuanced interpretations of object use and context that go beyond what conventional rule-based systems or object metadata can provide. Study 2 showed that these inferred properties, when used to drive haptic rendering, lead to measurable improvements in user experience—especially for spatial awareness, but also for perceived materiality and usability. Study 3 showed that the end-to-end pipeline successfully enhanced the overall user experience when the user is interacting in a full VR scene. Together, these results validate Scene2Hap's full pipeline: from automatic data extraction

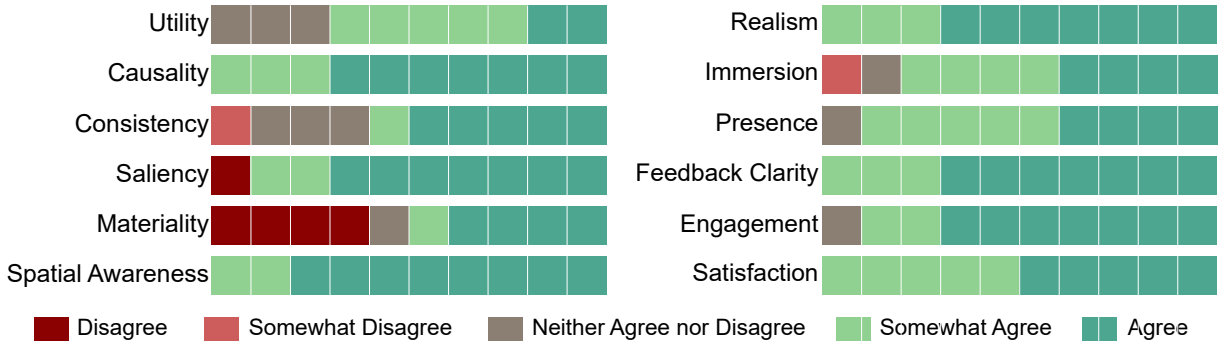


Figure 9: Results of Study 3. In this experiment, participants filled out two questionnaires. Questionnaire A contained items for four usability requirements (utility, causality, consistency, saliency) as well as perceived materiality and spatial awareness (left), and Questionnaire B included perceived realism, immersion, presence, feedback clarity, engagement, and satisfaction (right).

to LLM-based inference to physically grounded, perceptually meaningful haptic output.

A unique strength of Scene2Hap lies in its system architecture, which combines LLM-based inference with physical modeling. At its core, the system asks two distinct questions about each object in the scene: “How does it vibrate?” and “What are its material properties?”. The answer to the first question is used to retrieve or generate an audio file representing the object’s vibration. The answer to the second provides material properties such as density and stiffness, which are used in a physical model to determine how vibration propagates across connected surfaces. This model is then used to modulate and attenuate the live audio stream in real time, so that users feel the vibration that is appropriate to the location and material they are touching. This unique combination of semantic reasoning and real-time physical modeling enables Scene2Hap to generate haptic feedback that is adaptive, coherent, and requires no manual tuning.

This architecture provides practical benefits for VR designers. Scene2Hap enables rapid deployment of haptic experiences without requiring manual parameter tuning or specialized domain knowledge. It allows VR creators to build rich, multimodal environments at scale—even in scenes with many complex objects—making haptic feedback more accessible as a design material.

Scene2Hap has several limitations. First, object semantics are currently limited to scene-level use and binary vibration behavior; future work could support richer object states, part-level reasoning, or event-based triggering. Second, while our physical model supports real-time propagation, it assumes simplified geometries and the vibration propagation to only neighboring objects. It may benefit from higher-fidelity models if performance allows. Third, while audio quality is dependent on retrieval/generation methods, which are outside our scope, future work could provide quality control through advanced LLM-based selection strategies or integrate alternative automatic haptic generation methods (e.g., [77]) into the Scene2Hap pipeline. Fourth, the system’s performance is dependent on the specific LLM used (we used GPT-4o). While we mitigate potential non-deterministic outputs with a low temperature, we anticipate that future models will enhance accuracy and processing speed. This advancement may also resolve limitations regarding

input data, potentially achieving robust performance with fewer text or image inputs than currently used. Lastly, our approach was specifically designed for experiencing haptic vibrations that are triggered by mechanisms, machines, or other active sources in a VR scene. This approach can also be applied for a wider range of expressive VR scenes, such as rendering the floor-shaking resonance of a virtual music concert or rendering symbolic haptic feedback for magical effects. In future work, we hope to extend the approach to haptic experiences caused by user interaction, ranging from material properties, like friction or texture, to abstract feedback such as subtle UI confirmation pulses.

7 Conclusion

We present Scene2Hap, an LLM-centered system that automatically designs object-level vibrotactile feedback for entire VR scenes based on the objects’ semantic attributes and physical context. Scene2Hap comprises two main technical contributions: LLM-based haptic inference and physics-inspired haptic rendering. Scene2Hap performs LLM-based haptic inference that employs a multimodal large language model to estimate the semantics and physical context of each object, including its material properties and vibration behavior, from the multimodal information present in the VR scene. This semantic and physical context is then used to create plausible vibrotactile signals by generating or retrieving audio signals and converting them to vibrotactile signals. For the more realistic spatial rendering of haptics in VR, Scene2Hap performs physics-inspired haptic rendering in real-time that calculates the propagation and attenuation of vibration signals from their source across objects in the scene, considering the estimated material properties and physical contexts, such as the distance and contact between virtual objects. Results from three studies confirmed that (1) LLM-based haptic inference could successfully understand the semantics and physical contexts of various objects in VR scenes; (2) Physics-inspired haptic rendering significantly contributed to providing immersive VR haptic experiences by improving the sense of materiality and spatial awareness with plausible vibrotactile signals and vibration attenuation; (3) End-to-end pipeline successfully enhanced the overall user experience when interacting in a full VR scene.

Acknowledgments

Arata Jingu is a recipient of Google PhD Fellowship. This project received funding from the European Research Council (ERC, KinestheticDisplays, 101165100). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the funding organizations. Neither the funding organizations nor the granting authorities can be held responsible for them.

References

- [1] 9t5. 2025. *9t5 Low Poly House Bathroom & Laundry*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/props/9t5-low-poly-house-bathroom-laundry-166522>
- [2] Jan Achenbach. 2012. *Wave propagation in elastic solids*. Elsevier.
- [3] John Aitken. 1878. An account of some experiments on rigidity produced by centrifugal force. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 5, 29 (1878), 81–105.
- [4] Gerald Albaum, Roger Best, and Del Hawkins. 1981. Continuous vs discrete semantic differential rating scales. *Psychol. Rep.* 49, 1 (Aug. 1981), 83–86.
- [5] Easa AliAbbasi, Akbar Allahverdizadeh, Behnam Dadashzadeh, and Reza Jahangiri. 2020. Non-contact AC current measurement using vibration analysis of a MEMS piezoelectric cantilever beam. *Journal of Energy Management and Technology* 4, 4 (2020), 28–35.
- [6] Easa AliAbbasi, MReza Alipour Sormoli, and Cagatay Basdogan. 2022. Frequency-dependent behavior of electrostatic forces between human finger and touch screen under electroadhesion. *IEEE Transactions on Haptics* 15, 2 (2022), 416–428.
- [7] Muhammed Alzaim, Abdulgazi Gedik, and Abdullah Hilmi Lav. 2020. Effect of modulus of bituminous layers and utilization of capping layer on weak pavement subgrades. *Civil Engineering Journal* 6, 7 (2020), 1286–1299.
- [8] AmesWeb. 2023. Young's Modulus of Steel. <https://amesweb.info/Materials/Youngs-Modulus-of-Steel.aspx> Accessed: 2025-04-09.
- [9] M Arndt, RD Machado, and A Scremin. 2010. An adaptive generalized finite element method applied to free vibration analysis of straight bars and trusses. *Journal of Sound and Vibration* 329, 6 (2010), 659–672.
- [10] Yuki Ban and Yusuke Ujitoko. 2018. TactGAN: Vibrotactile Designing Driven by GAN-based Automatic Generation. In *SIGGRAPH Asia 2018 Emerging Technologies (SA '18)*. Association for Computing Machinery, New York, NY, USA, 1–2. doi:10.1145/3275476.3275484
- [11] Gloria Beltrán and Miguel Romo. 2012. Analysis of Poisson's Ratio Effect on Pavement Layer Moduli Estimation-A Neural Network Based Approach from Non-destructive Testing. In *Ibero-American Conference on Artificial Intelligence*. Springer, 371–380.
- [12] GoN Bycroft. 1956. Forced vibrations of a rigid circular plate on a semi-infinite elastic space and on an elastic stratum. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 248, 948 (1956), 327–368.
- [13] Shaoyu Cai, Lu Zhao, Yuki Ban, Takuji Narumi, Yue Liu, and Kening Zhu. 2022. GAN-based image-to-friction generation for tactile simulation of fabric material. *Computers & Graphics* 102 (Feb. 2022), 460–473. doi:10.1016/j.cag.2021.09.007
- [14] Shaoyu Cai and Kening Zhu. 2022. Multi-modal Transformer-based Tactile Signal Generation for Haptic Texture Simulation of Materials in Virtual and Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 810–811. doi:10.1109/ISMAR-Adjunct57072.2022.00174
- [15] Shaoyu Cai, Kening Zhu, Yuki Ban, and Takuji Narumi. 2021. Visual-Tactile Cross-Modal Data Generation Using Residue-Fusion GAN With Feature-Matching and Perceptual Losses. *IEEE Robotics and Automation Letters* 6, 4 (Oct. 2021), 7525–7532. doi:10.1109/LRA.2021.3095925
- [16] Guanqun Cao, Jiaqi Jiang, Ningtao Mao, Danushka Bollegala, Min Li, and Shan Luo. 2023. Vis2Hap: Vision-based Haptic Rendering by Cross-modal Generation. doi:10.48550/arXiv.2301.06826
- [17] Li-Qun Chen. 2005. Analysis and control of transverse vibrations of axially moving strings. *Appl. Mech. Rev.* 58, 2 (2005), 91–116.
- [18] SM Cramer, OM Friday, RH White, and G Sriprutkiat. 2003. Mechanical properties of gypsum board at elevated temperatures. In *Fire and materials 2003: 8th International Conference* (San Francisco, CA, USA). London: Interscience Communications Limited, 33–42.
- [19] Lothar Cremer and Manfred Heckl. 2013. *Structure-borne sound: structural vibrations and sound radiation at audio frequencies*. Springer Science & Business Media.
- [20] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time Prompting of Interactive Worlds using Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3642579
- [21] Davide Deflorio, Massimiliano Di Luca, and Alan M Wing. 2022. Skin and mechanoreceptor contribution to tactile input for perception: A review of simulation models. *Frontiers in Human Neuroscience* 16 (2022), 862344.
- [22] Donald Degraen, Bruno Fruchard, Frederik Smolders, Emmanouil Potetsianakis, Seref Güngör, Antonio Krüger, and Jürgen Steimle. 2021. Weiriding Haptics: In-Situ Prototyping of Vibrotactile Feedback in Virtual Reality through Vocalization. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 936–953. doi:10.1145/3472749.3474797
- [23] Haiwei Dong, Yu Gao, Hussein Al Osman, and Abdulmoteleb El Saddik. 2015. Development of a Web-Based Haptic Authoring Tool for Multimedia Applications. In *2015 IEEE International Symposium on Multimedia (ISM)*. 13–20. doi:10.1109/ISM.2015.71
- [24] CES Edupack. 2009. Plywood (5 ply, beech), parallel to face layer. <https://simulatentoast.wordpress.com/wp-content/uploads/2013/03/plywood.pdf> Accessed: 2025-04-08.
- [25] Alper Erturk and Daniel J Inman. 2009. An experimentally validated bimorph cantilever model for piezoelectric energy harvesting from base excitations. *Smart materials and structures* 18, 2 (2009), 025009.
- [26] Faraz Faruqi, Maxine Perroni-Scharf, Jaskaran Singh Walia, Yunyi Zhu, Shuyue Feng, Donald Degraen, and Stefanie Mueller. 2025. TactStyle: Generating Tactile Textures with Generative AI for Digital Fabrication. doi:10.1145/3706598.3713740 arXiv:2503.02007 [cs].
- [27] Alexander Fluegel. 2007. Calculation of Poisson's Ratio of Glasses. https://www.glassproperties.com/poisson_ratio/ Accessed: 2025-04-09.
- [28] Zayd Hajje. 2024. On the exponential decay of a Balakrishnan-Taylor plate with strong damping. *AIMS Mathematics* 9, 6 (2024), 14026–14042.
- [29] Negin Heravi, Heather Culbertson, Allison M. Okamura, and Jeannette Bohg. 2023. Development and Evaluation of a Learning-based Model for Real-time Haptic Texture Rendering. doi:10.48550/arXiv.2212.13332
- [30] Rebekka Hoffmann, Simone Spagnol, Árni Kristjánsson, and Runar Unnthorsson. 2018. Evaluation of an audio-haptic sensory substitution device for enhancing spatial awareness for the visually impaired. *Optometry and Vision Science* 95, 9 (2018), 757–765.
- [31] Guimin Hu, Daniel Hershovich, and Hasti Seifi. 2025. HapticCap: A Multimodal Dataset and Task for Understanding User Experience of Vibration Haptic Signals. doi:10.48550/arXiv.2507.13318 arXiv:2507.13318 [cs].
- [32] Guimin Hu, Daniel Hershovich, and Hasti Seifi. 2025. HapticLLaMa: A Multimodal Sensory Language Model for Haptic Captioning. doi:10.48550/arXiv.2508.06475 arXiv:2508.06475 [cs].
- [33] Lukas Höllein, Ang Cao, Andrew Owens, Justin Johnson, and Matthias Nießner. 2023. Text2Room: Extracting Textured 3D Meshes from 2D Text-to-Image Models. 7909–7920. https://openaccess.thecvf.com/content/ICCV2023/html/Hollein_Text2Room_Extracting_Textured_3D_Meshes_from_2D_Text-to-Image_Models_ICCV_2023_paper.html
- [34] Arkema Inc. 2002. Plexiglas G Acrylic Sheet. https://www.alro.com/dataPDF/Plastics/PlasticsBrochures/Brochure_PlexiglasG.pdf Accessed: 2025-04-08.
- [35] Altra Infinite. 2025. *Kitchen Appliance - Low Poly*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/props/electronics/kitchen-appliance-low-poly-180419>
- [36] Ali Israr and Ivan Poupyrev. 2011. Tactile brush: drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2019–2028. doi:10.1145/1978942.1979235
- [37] Ali Israr, Siyan Zhao, Kaitlyn Schwalje, Roberta Klatzky, and Jill Lehman. 2014. Feel Effects: Enriching Storytelling with Haptic Feedback. *ACM Trans. Appl. Percept.* 11, 3 (Sept. 2014), 11:1–11:17. doi:10.1145/2641570
- [38] Kevin John, Yinan Li, and Hasti Seifi. 2024. AdapTics: A Toolkit for Creative Design and Integration of Real-Time Adaptive Mid-Air Ultrasound Tactons. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3613904.3642090
- [39] David Vaughan Jones. 1987. *The surface propagation of ground vibration*. Ph.D. Dissertation. University of Southampton.
- [40] Jakob Kilian, Alexander Neugebauer, Lasse Scherffig, and Siegfried Wahl. 2022. The unfolding space glove: A wearable spatio-visual to haptic sensory substitution device for blind people. *Sensors* 22, 5 (2022), 1859.
- [41] Dong-Geun Kim, Jungeun Lee, Gyeore Yun, Hong Z. Tan, and Seungmoon Choi. 2024. Sound-to-Touch Crossmodal Pitch Matching for Short Sounds. *IEEE Transactions on Haptics* 17, 1 (Jan. 2024), 2–7. doi:10.1109/TOH.2023.3338224
- [42] Erin Kim and Oliver Schneider. 2020. Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376280
- [43] Erin Kim and Oliver Schneider. 2020. Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI,*

- USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376280
- [44] Leonidas Alexandros S Kouris, Dionysios A Bournas, Olufemi T Akintayo, Avraam A Konstantinidis, and Elias C Aifantis. 2020. A gradient elastic homogenisation model for brick masonry. *Engineering Structures* 208 (2020), 110311.
- [45] Felix Kreuk, Gabriel Synnaeve, Adam Polyak, Uriel Singer, Alexandre Défossez, Jade Copet, Devi Parikh, Yaniv Taigman, and Yossi Adi. 2023. AudioGen: Textually Guided Audio Generation. doi:10.48550/arXiv.2209.15352
- [46] Jaebong Lee and Seungmoon Choi. 2013. Real-time perception-level translation from audio signals to vibrotactile effects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2567–2576. doi:10.1145/2470654.2481354
- [47] Arthur W Leissa. 1973. *Vibration of shells*. Vol. 288. NASA SP-288, National Aeronautics and Space Administration.
- [48] Xinwu Li, Huaping Liu, Junfeng Zhou, and FuChun Sun. 2019. Learning cross-modal visual-tactile representation using ensembled generative adversarial networks. *Cognitive Computation and Systems* 1, 2 (2019), 40–44. doi:10.1049/ccs.2018.0014
- [49] Chungman Lim, Kevin John, Jin Gyungmin, Hasti Seifi, and Gunhyuk Park. 2025. ChatHAP: A Chat-Based Haptic System for Designing Vibrations through Conversation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, Yokohama, Japan.
- [50] Augustus Edward Hough Love. 1929. IX. The stress produced in a semi-infinite solid by pressure on part of the boundary. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 228, 659–669 (1929), 377–420.
- [51] AZO Materials. 2025. An Introduction to Copper. <https://www.azom.com/properties.aspx?ArticleID=597> Accessed: 2025-04-09.
- [52] AZO Materials. 2025. What Are the Properties of Aluminum? <https://www.azom.com/properties.aspx?ArticleID=1446> Accessed: 2025-04-09.
- [53] Cristina Mayo-Corrochano, Luis Javier Sánchez-Aparicio, José-Ramón Aira, David Sanz-Arauz, Esther Moreno, and Javier Pinilla Melo. 2022. Assessment of the elastic properties of high-fired gypsum using the digital image correlation method. *Construction and Building Materials* 317 (2022), 125945.
- [54] Glowing Moth. 2025. *Construction Props*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/props/tools/construction-props-161383>
- [55] Dimmukhamed Mukashev, Nimesha Ranasinghe, and Aditya Shekhar Nittala. 2023. TactTongue: Prototyping ElectroTactile Stimulation on the Tongue. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3586183.3606829
- [56] Momoka Nakayama, Risako Kawashima, Shintaro Murakami, Yuta Takeuchi, Tatsuya Mori, and Dai Takanashi. 2024. A Method for Generating Tactile Sensations from Textual Descriptions Using Generative AI. In *SIGGRAPH Asia 2024 Posters (SA '24)*. Association for Computing Machinery, New York, NY, USA, 1–2. doi:10.1145/3681756.3697870
- [57] Seoyong Nam, Minh Chung, Haerim Kim, Eunhae Kim, Taehyeon Kim, and Yongjae Yoo. 2024. Automatic Generation of Multimodal 4D Effects for Immersive Video Watching Experiences. In *SIGGRAPH Asia 2024 Technical Communications (SA '24)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3681758.3698021
- [58] JM Nichols and YZ Toteov. 1997. Experimental determination of the dynamic Modulus of Elasticity of masonry units. In *15th Australian Conference on the Mechanics of Structures and Materials* (Melbourn, Vic, Australia). 1–7.
- [59] nickknacks. 2025. *Garden Essentials*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/props/garden-essentials-64116>
- [60] Ryuta Okazaki, Hidenori Kuribayashi, and Hiroyuki Kajimoto. 2015. The Effect of Frequency Shifting on Audio–Tactile Conversion for Enriching Musical Experience. In *Haptic Interaction: Perception, Devices and Applications*, Hiroyuki Kajimoto, Hideyuki Ando, and Ki-Uk Kyung (Eds.). Springer Japan, Tokyo, 45–51. https://doi.org/10.1007/978-4-431-55690-9_9
- [61] A Ranjbaran, AR Shokrzadeh, and S Khosravi. 2011. A new finite element analysis of free axial vibration of cracked bars. *International Journal for Numerical Methods in Biomedical Engineering* 27, 10 (2011), 1611–1621.
- [62] Singiresu S Rao. 2019. *Vibration of continuous systems*. John Wiley & Sons.
- [63] Lord Rayleigh. 1885. On waves propagated along the plane surface of an elastic solid. *Proceedings of the London Mathematical Society* 1, 1 (1885), 4–11.
- [64] Qiaoqiao Ren and Tony Belpaeme. 2025. Touched by ChatGPT: Using an LLM to Drive Affective Tactile Interaction. doi:10.48550/arXiv.2501.07224
- [65] Oliver Schneider, Karon MacLean, Colin Swindells, and Kellogg Booth. 2017. Haptic experience design: What hapticians do and where they need help. *International Journal of Human-Computer Studies* 107 (Nov. 2017), 5–21. doi:10.1016/j.ijhcs.2017.04.004
- [66] Oliver Schneider, Siyan Zhao, and Ali Israr. 2015. FeelCraft: User-Crafted Tactile Content. In *Haptic Interaction: Perception, Devices and Applications*, Hiroyuki Kajimoto, Hideyuki Ando, and Ki-Uk Kyung (Eds.). Springer Japan, Tokyo, 253–259. https://doi.org/10.1007/978-4-431-55690-9_47
- [67] Oliver S. Schneider, Ali Israr, and Karon E. MacLean. 2015. Tactile Animation by Direct Manipulation of Grid Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Association for Computing Machinery, New York, NY, USA, 21–30. doi:10.1145/2807442.2807470
- [68] Inc. SCHOTT North America. 2004. TIE-31: Mechanical and thermal properties of optical glass. https://wp.optics.arizona.edu/optomech/wp-content/uploads/sites/53/2016/10/tie-31_mechanical_and_thermal_properties_of_optical_glass_us.pdf Accessed: 2025-04-09.
- [69] Hasti Seifi, Sean Chew, Antony James Nascé, William Edward Lowther, William Frier, and Kasper Hornbæk. 2023. Feellustrator: A Design Tool for Ultrasound Mid-Air Haptics. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3580728
- [70] Hasti Seifi, Matthew Chun, Colin Gallacher, Oliver Schneider, and Karon E. MacLean. 2020. How Do Novice Hapticians Design? A Case Study in Creating Haptic Learning Environments. *IEEE Transactions on Haptics* 13, 4 (Oct. 2020), 791–805. doi:10.1109/TOH.2020.2968903
- [71] Sedanur Seker and Kucuk Huseyin Koc. 2023. Engineering Analysis Application in Furniture Making: Deformation-Equivalent Stress. *Drewno: Prace Naukowe, Doniesienia, Komunikaty* 66, 211 (2023), Art. no. 1644–3985.406.07.
- [72] Uriel Singer, Shelly Sheynin, Adam Polyak, Oron Ashual, Iurii Makarov, Filippos Kokkinos, Naman Goyal, Andrea Vedaldi, Devi Parikh, Justin Johnson, and Yaniv Taigman. 2023. Text-to-4D dynamic scene generation. In *Proceedings of the 40th International Conference on Machine Learning (ICML '23, Vol. 202)*. JMLR.org, Honolulu, Hawaii, USA, 31915–31929.
- [73] Maciej Stroinski, Kamil Kwarciak, Mateusz Kowalewski, Daria Hemmerling, William Frier, and Orestis Georgiou. 2024. Text-to-Haptics: Enhancing Multi-sensory Storytelling through Emotionally Congruent Midair Haptics. *Advanced Intelligent Systems* n/a, n/a (2024), 2400758. doi:10.1002/aisy.202400758
- [74] Little Arms Studios. 2025. *Painless Props - Bathroom and Laundry*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/painless-props-bathroom-and-laundry-43042>
- [75] Star Game Studios. 2025. *Stylized Kitchen Furniture*. Retrieved April 19, 2025 from <https://assetstore.unity.com/packages/3d/props/furniture/stylized-kitchen-furniture-189711>
- [76] Xia Su, Jon E. Froehlich, Eunye Koh, and Chang Xiao. 2024. SonifyAR: Context-Aware Sound Generation in Augmented Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3654777.3676406
- [77] Youjin Sung, Kevin John, Sang Ho Yoon, and Hasti Seifi. 2025. HapticGen: Generative Text-to-Vibration Model for Streamlining Haptic Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.
- [78] Youjin Sung, Rachel Kim, Kun Woo Song, Yitian Shao, and Sang Ho Yoon. 2024. HapticPilot: Authoring In-situ Hand Posture-Adaptive Vibrotactile Feedback for Virtual Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4 (Jan. 2024), 179:1–179:28. doi:10.1145/3631453
- [79] Stephen Timoshenko. 1983. *History of strength of materials: with a brief account of the history of theory of elasticity and theory of structures*. Courier Corporation.
- [80] Stephen Timoshenko and Sergius Woinowsky-Krieger. 1959. Theory of plates and shells. (1959).
- [81] The Engineering ToolBox. 2011. Wood, Panel and Structural Timber Products - Mechanical Properties. https://www.engineeringtoolbox.com/timber-mechanical-properties-d_1789.html Accessed: 2025-04-09.
- [82] TC Tsai and YZ Wang. 1996. Vibration analysis and diagnosis of a cracked shaft. *Journal of Sound and Vibration* 192, 3 (1996), 607–620.
- [83] Yusuke Ujitoko and Yuki Ban. 2018. Vibrotactile Signal Generation from Texture Images or Attributes Using Generative Adversarial Network. In *Haptics: Science, Technology, and Applications (Lecture Notes in Computer Science)*, Domenico Prattichizzo, Hiroyuki Shinoda, Hong Z. Tan, Emanuele Ruffaldi, and Antonio Frisoli (Eds.). Springer International Publishing, Cham, 25–36. doi:10.1007/978-3-319-93399-3_3
- [84] Yusuke Ujitoko, Yuki Ban, and Koichi Hirota. 2020. GAN-Based Fine-Tuning of Vibrotactile Signals to Render Material Surfaces. *IEEE Access* 8 (2020), 16656–16661. doi:10.1109/ACCESS.2020.2968185
- [85] Anke van Oosterhout, Miguel Bruns, and Eve Hoggan. 2020. Facilitating Flexible Force Feedback Design with Felix. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20)*. Association for Computing Machinery, New York, NY, USA, 184–193. doi:10.1145/3382507.3418819
- [86] Iddo Yehoshua Wald, Donald Degraen, Amber Maimon, Jonas Keppel, and Stefan Schneegass. 2025. Spatial Haptics: A Sensory Substitution Method for Distal Object Detection Using Tactile Cues. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, Yokohama, Japan.
- [87] Wikipedia. 2025. Poly(methyl methacrylate). [https://en.wikipedia.org/wiki/Poly\(methyl_methacrylate\)#cite_note-p2-16](https://en.wikipedia.org/wiki/Poly(methyl_methacrylate)#cite_note-p2-16) Accessed: 2025-04-09.
- [88] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova

procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (CHI '11). Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

- [89] Haoyu Zhao, Hao Wang, Xingyue Zhao, Hongqiu Wang, Zhiyu Wu, Chengjiang Long, and Hua Zou. 2025. Efficient Physics Simulation for 3D Scenes via MLLM-Guided Gaussian Splatting. arXiv:2411.12789 <https://arxiv.org/abs/2411.12789>

A Used Prompt Templates

This appendix section introduces the prompt templates used for our LLM components in Study 1. The bracketed sections in these templates are automatically replaced with the information for each scene or object.

A.1 Initial Simple Prompt

```
1 Your role is (1) to recognize the contexts
  of a Unity gameobject from its name,
  size, position, and images, (2) to
  estimate the material properties, (3)
  to describe how an object should
  vibrate in a Unity scene.
2
3 The name of the Unity scene is {scene_name}.
4 The sent images comprise three sets. The
  first {len_scene} images sent were
  taken from different angles in the
  scene. The next {len_isolated} images
  are isolated images that show an object
  of interest in the center part from
  different angles. The other
  {len_context} images are scene images
  that show the same object in the scene
  from different angles.
5 The user prompt is {user_prompt}.
6 The object name in a Unity scene is
  {object_name}.
7 The size of the object in the scene is
  {size} in a meter unit.
8 The object is placed at the Y position of
  {position_y} in the scene in a meter
  unit.
9
10 Estimate its actual size in a string format
  like '1.0,1.0,1.0'.
11 Estimate whether the object should vibrate
  in the scene in some cases (bool).
12
13 Estimate its density in kg/m^3, Young's
  modulus in GPa, Poisson's ratio, and
  damping ratio of the material category
  in float values.
14
15 If the object should vibrate, answer the
  following. If the object should not
  vibrate, return an empty string.
16 Describe how the object should vibrate with
  less than 15 words.
```

```
17 In Addition, provide keywords that describe
  the vibration by connecting two sets of
  words with blanks like '<Keyword A>
  <Keyword B>'.
```

```
18 Provide the estimated size and its reason,
  whether the object should vibrate and
  its reason, density, Young's modulus,
  Poisson's ratio, damping ratio,
  free-form vibration description, and
  keywords in a JSON format without any
  affixes. All structured outputs should
  be provided.
```

A.2 Final Prompt for Scene Analyzer

```
1 Your role is to recognize the category of a
  Unity scene from its name and images.
2 The name of the Unity scene is {scene_name}.
3 The images sent were taken from different
  angles in the scene.
4
5 Estimate its scene category in 1-2 words
  from its name and images.
6 This category should be very specific
  without ambiguity. {scene_name} does
  not necessarily mean the correct scene
  category.
7 The scene category should be the name of
  its environment or scene, not a summary
  of the objects in images.
8 Take into account only the images showing
  objects clearly, and ignore the other
  images.
9
10 Provide the scene category without any
  affixes. If it is extremely difficult
  to estimate the scene category, answer
  'undefined'.
```

A.3 Final Prompt for Object Analyzer

```
1 Your role is to recognize the contexts of a
  Unity gameobject from its name, size,
  position, and images.
2 The user prompt is {user_prompt}. If the
  user prompt is not empty, conduct the
  below estimation with the highest
  importance on the user prompt.
3 The scene category of the Unity scene is
  {scene_category}.
4 The object name in a Unity scene is
  {object_name}.
```

5 The size of the object in the scene is {size} in a meter unit. It is not decided which value of this size vector is the width, height, or depth. This size is a dimension of the dominant surface of the object. For example, if the object is a table with legs, the value is the size of the tabletop.

6 The object is placed at the Y position of {position_y} in the scene in a meter unit.

7 The sent images comprise two sets. The first {len_isolated} images are isolated images that show an object of interest in the center part from different angles. The other {len_scene} images are scene images that show the same object in the scene from different angles.

8

9 Estimate its object category in 1-3 words from its name, size, position, and images. However, if {object_name} sounds like a boundary surface (e.g., floor, ceiling, wall) or a room, give the most importance for estimation to its object name and ignore its size.

10 This object category should be very specific without ambiguity (e.g., 'refrigerator' is better than 'appliance' in terms of clarity). {object_name} is not necessarily the correct object category.

11 If there are multiple options for the object category, choose the one that is most likely to exist in {scene_category}. Try not to choose a category that is not likely to exist in {scene_category}.

12 When you check the scene images, estimate the object category of only the object surrounded in a pink outline, and not consider the whole environment. If this pink outline does not completely surround an object or is not visualized at all in the scene images, consider the target object to be the object in the center of the scene images and most resembles the object in the isolated images.

13 Take into account only the images showing some objects clearly, and ignore the other images.

14 Take into account the object's authenticity based on whether it is being used in a physically plausible way in the scene images and whether its size roughly matches the typical size of its object category that humans use in everyday environments. This size check should not be too strict. If this object is not authentic, include a word to describe the authenticity (e.g., 'miniature' if the object is too small) in the estimated object category.

15 Position information can be used to estimate the object category, especially it has an ambiguous name and shape.

16 Do not estimate the object category from the light and reflective conditions because the images are taken from various lighting conditions.

17

18 If the object is a boundary surface, it is likely that one axis of {size} is too small in Unity. In that case, estimate the object size by replacing only that axis value with a typical value for the object category in meters and provide a reason in one sentence. Return the same value as {size} for the estimated size in the other cases. Note that you should return the value in a string format like '1.0,1.0,1.0'. For example, if the thickness of the room floor is too small, replace it with a typical value for the room floor.

19 Estimate its material category in 1 word from its isolated images and object category. If the object comprises multiple materials, choose the most dominant material. This material category should be as specific as possible, not a general term. (e.g., 'iron' or 'steel' should be used rather than 'metal' in terms of concreteness). If the object is not authentic, estimate the material category based on the object's authenticity. If the object seems a boundary surface and is textureless, estimate the material that is likely to be present in the {scene_category} based on its surface color.

20 Estimate how the object should be used in the scene in one sentence from the scene images. If humans generally use the object while holding it in the scene, consider that case.

Estimate whether the object should vibrate in the scene in some cases (bool) based on its scene images and estimated usage. For example, the target object could vibrate due to thermal energy propagated from surrounding objects or its internal mechanism. If humans generally use the object while holding it in the scene, consider that case. If the target object or an adjacent object is an electric machine, consider the vibration that can occur when they are powered on. Do not consider the propagation of mechanical vibration originating from adjacent objects.

Provide the object category and its reason, material category, usage, estimated size and its reason, whether the object should vibrate and its reason in a JSON format without any affixes. All structured outputs should be provided.

A.4 Final Prompt for Material Property Estimator

Your role is to estimate the material properties of a material category. Estimate density in kg/m^3 , Young's modulus in GPa, Poisson's ratio, and damping ratio of {material_category} in float values. Strictly check that the values are provided in the correct unit. Provide these numerical values in a JSON format without any affixes or units. All structured outputs should be provided. If you cannot estimate the material properties for some reason, assign 0 for all values.

A.5 Final Prompt for Vibration Describer

Your role is to describe how an object should vibrate in a Unity scene. {object_category} is used in the following way: {usage}. Describe how the object should vibrate in a simple and straightforward sentence with less than 15 words. This sentence should start from {object_category} and mention its vibration characteristics in simple words.

In Addition, provide keywords that describe the vibration by connecting two sets of words with blanks like '<Keyword A> <Keyword B>'. The first set has to be {object_category}. The second keyword should be one verb in its base form related to the vibration that best describes how the object vibrates in the scene. Do not use the word 'vibrate' in the keywords.

Provide the free-form sentence and the combined keywords in a JSON format without any affixes. All structured outputs should be provided.

B Study 2 Materials

B.1 Questions

We used the following questions to evaluate vibration propagation. Each question was accompanied by a continuous line, anchored with “Strongly Disagree” on the left and “Strongly Agree” on the right. Participants marked a point along the line, which was then measured and recorded as a percentage.

- **Utility:** Haptic feedback was able to benefit my user experience in a way that other sensory modalities cannot.
- **Causality:** I could identify and describe the source of haptic feedback.
- **Consistency:** The system's ability to generate the proper haptic feedback was reliable.
- **Saliency:** The noticeability of haptic feedback was correct as it related to its purpose and context.
- **Materiality:** Haptic feedback helped me understand the type of material that I touched.
- **Spatial Awareness:** Haptic feedback helped me better perceive the virtual space.

B.2 Data by Scene

For the sake of completeness, we provide an overview of the recorded data by scene in Figure 10. To ensure the validity of treating scenes as repetitions in the analysis provided in subsection 5.2, we conducted a repeated measures ANOVA on each item to identify if there were significant differences between scenes, and found none.

- **Utility:** $F(2, 18) = 0.579, p = .571$.
- **Causality:** $F(2, 18) = 1.501, p = .250$.
- **Consistency:** $F(2, 18) = 1.421, p = .267$.
- **Saliency:** $F(2, 18) = 0.586, p = .567$.
- **Materiality:** $F(2, 18) = 1.656, p = .219$.
- **Spatial Awareness:** $F(2, 18) = 1.968, p = .169$.

C Study 3 Materials

C.1 Questions

We used the following questions to evaluate the full-VR scene. Each question was accompanied by a five-point Likert scale, anchored with the following options from left to right: “Disagree”, “Somewhat

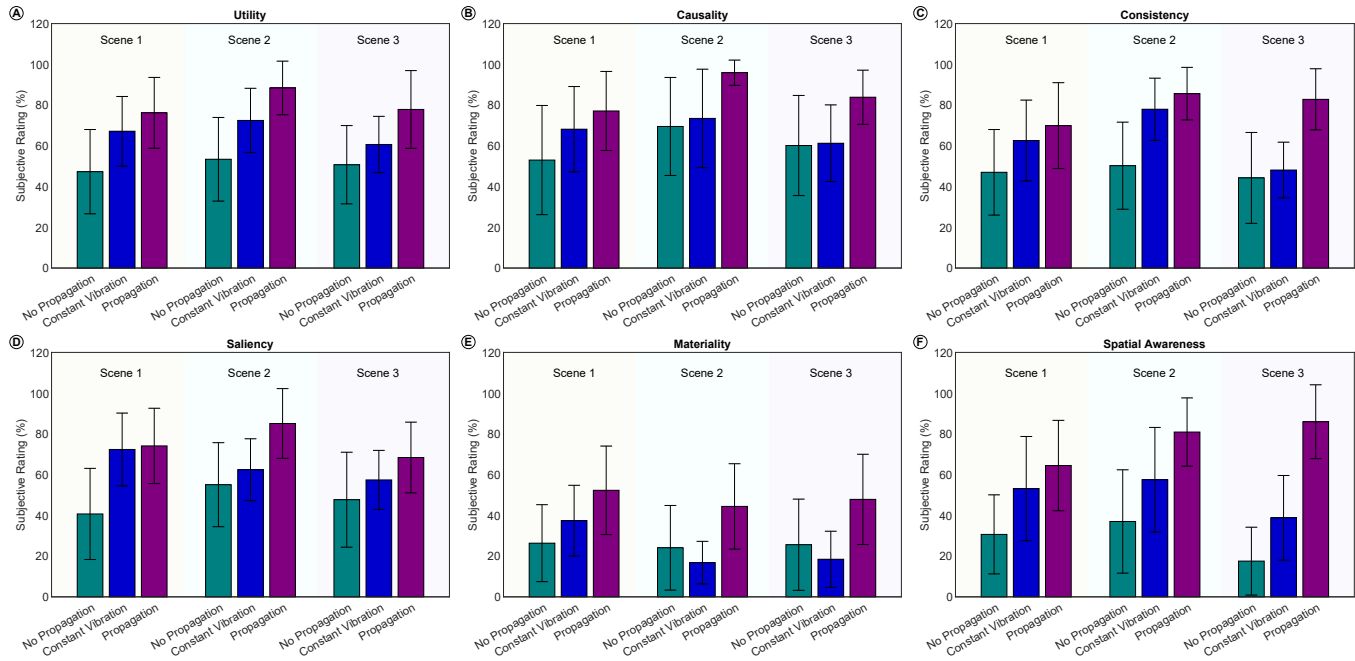


Figure 10: Overview of means and confidence intervals, broken down by scene

Disagree”, “Neither Agree nor Disagree”, “Somewhat Agree”, and “Agree”.

- **Realism:** Haptic Feedback enhanced the realism of the scene.
- **Immersion:** Haptic Feedback enhanced my immersion in the scene.
- **Presence:** Haptic Feedback enhanced my presence in the virtual environment – as if I was really there.

- **Feedback Clarity:** It was easier to understand interactions with haptic feedback.
- **Engagement:** I felt more engaged with the environment because of haptic feedback.
- **Satisfaction:** The experience was more enjoyable because of haptic feedback.