HapticGen: Generative Text-to-Vibration Model for Streamlining Haptic Design

Youjin Sung*
Graduate School of Culture Technology
KAIST
Daejeon, Republic of Korea
672@kaist.ac.kr

Kevin John*
School of Computing and Augmented Intelligence
Arizona State University
Tempe, Arizona, USA
kevin.john@asu.edu

Sang Ho Yoon†
Graduate School of Culture Technology
KAIST
Daejeon, Republic of Korea
sangho@kaist.ac.kr

Hasti Seifi†
School of Computing and Augmented Intelligence
Arizona State University
Tempe, Arizona, USA
Department of Computer Science
University of Copenhagen
Copenhagen, Denmark
hasti.seifi@asu.edu

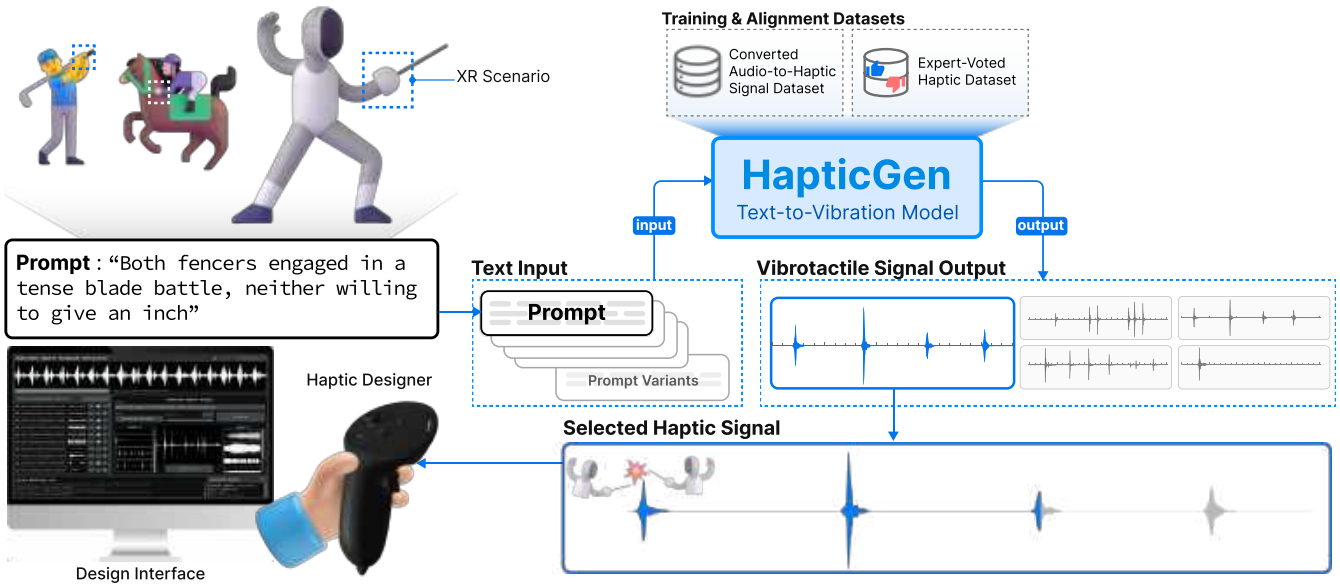


Figure 1: Overview of HapticGen. Designers can enter a text prompt into HapticGen about an XR scenario (e.g., riding a horse or fencing). HapticGen then creates multiple prompt variants, generates vibration signals for the prompts, and presents output vibrations on its desktop-based user interface. The designers can select and play the vibrations on VR controllers, vote the signals with thumbs up or down, and revise the prompt to generate new vibrations or use the signals in their XR applications.

Abstract

Designing haptic effects is a complex, time-consuming process requiring specialized skills and tools. To support haptic design, we introduce HapticGen, a generative model designed to create vibrotactile signals from text inputs. We conducted a formative workshop to identify requirements for an AI-driven haptic model. Given the limited size of existing haptic datasets, we trained HapticGen on a large, labeled dataset of 335k audio samples using an automated audio-to-haptic conversion method. Expert haptic designers then used HapticGen's integrated interface to prompt

*Youjin Sung and Kevin John contributed equally to the paper.
†Sang Ho Yoon and Hasti Seifi are the corresponding authors.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713609

and rate signals, creating a haptic-specific preference dataset for fine-tuning. We evaluated the fine-tuned HapticGen with 32 users, qualitatively and quantitatively, in an A/B comparison against a baseline text-to-audio model with audio-to-haptic conversion. Results show significant improvements in five haptic experience (e.g., realism) and system usability factors (e.g., future use). Qualitative feedback indicates HapticGen streamlines the ideation process for designers and helps generate diverse, nuanced vibrations.

CCS Concepts

• **Human-centered computing** → **Haptic devices**; **Virtual reality**; • **Hardware** → *Tactile and hand-based interfaces*; • **Computing methodologies** → **Model development and analysis**.

Keywords

Haptics, Designers, Generative AI, Extended Reality

ACM Reference Format:

Youjin Sung, Kevin John, Sang Ho Yoon, and Hasti Seifi. 2025. HapticGen: Generative Text-to-Vibration Model for Streamlining Haptic Design. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3706598.3713609

1 Introduction

Vibrotactile feedback can enhance user experience in various applications, including virtual reality (VR), gaming, and assistive technologies [13, 36, 73, 86]. However, designing haptic feedback requires specialized skills and tools and is often a time-consuming process [63, 69]. Haptic designers face challenges in ideating and creating signals that capture the nuances and richness of the physical world or effectively convey the designer's meaning and intent to users. Moreover, evaluating haptic signals is often subjective and relies on the designer's intuition or user studies which further complicates the task for novice designers [37].

Generative AI (GenAI) has revolutionized the design process for other modalities such as text [15], images [16, 72], and audio [6, 39] by lowering skill barriers, supporting natural language controls, and reducing iteration time. Despite these advances, no generative text-to-haptic model currently exists [25]. One key barrier in haptics is the lack of large and diverse datasets for training such models [39, 60]. Existing libraries of haptic signals, such as VibViz [71], contain only a few hundred signals and a limited number of user ratings and tags. Collecting haptic data is time-consuming and costly, further limiting the development of data-driven haptic design tools. The difficulty of data collection is due to factors such as the lack of methods for creating diverse haptic signals at scale and the absence of a standard vocabulary for labeling signals [37, 45]. Haptic signals are also particularly difficult to label in isolation, as their interpretation often depends on contextual information, such as how well they align with a designer-provided description or goal, and their interaction with other stimuli in an application [45].

To support vibrotactile design, we introduce HapticGen, a generative model that can create vibrotactile haptic signals from textual inputs (Figure 1). HapticGen leverages recent advancements in generative text-to-audio models and integrates domain-specific modifications to optimize for haptic data generation and align with haptic designer preferences. Specifically, we used an autoregressive transformer model adapted from the architecture introduced in MusicGen [6] and retrained the EnCodec tokenizer for haptic signals [12]. We built an initial haptic dataset to train our model by applying automated filtering, haptic label augmentation, and audio-to-haptic conversion steps onto WavCaps [47], a large-scale dataset of ~400k audio samples. To further align the model for haptics, we collected a new text+vibration dataset by having haptics experts generate vibrations using the initial version of HapticGen and vote on their haptic quality. HapticGen was further refined on this expert-voted haptic dataset using fine-tuning techniques such as Direct Preference Optimization (DPO) [57]. We developed an interface for HapticGen that allows users to seamlessly generate, play, and rate vibrotactile signals using Meta Quest controllers [49, 51], chosen for their practicality as widely available commodity hardware, without requiring the use of the head mounted display (HMD).

We designed and evaluated HapticGen with three studies. First, we conducted a formative study with 9 haptics researchers to identify the unique requirements for an AI-driven haptic design tool. Second, we ran a design study with the initial version of HapticGen trained on haptic-converted WavCaps dataset, where 15 expert haptic designers provided prompts and rated vibration signals generated by the initial model. From this study, we collected the expert-voted haptic dataset to fine-tune the model and collected qualitative insights to improve the interface and model output. These improvements include automatically creating prompt variations to increase the diversity of haptic generation and signal normalization to avoid low-intensity vibration outputs. Finally, we evaluated the HapticGen text-to-vibration model in a comprehensive study with 32 participants, comparing its performance in generating vibration signals to a baseline text-to-audio generation model with automatic audio-to-vibration conversion. Results from this A/B comparison showed significant improvements ($p < 0.05$) in both the perceived haptic experience (Autotelics, Realism) and system usability (Workload, Future Use, Goal). The results also showed improved ratings for other factors (Expressivity, Iteration) compared to the baseline. Additionally, qualitative feedback from users highlighted that HapticGen streamlined the ideation process, enabled the creation of dynamic and nuanced vibrations, and made haptic design more accessible to novice designers.

The main contributions of this work are:

- HapticGen, a first generative model capable of creating diverse vibrotactile haptic signals from textual input.
- Two large captioned haptic datasets: 1) an expert-voted preference dataset that we used for fine-tuning the HapticGen model with 1297 tuples of [text prompt, vibration, thumbs up/down vote], and 2) a user-voted preference dataset from our final A/B testing study with 3229 tuples (see Appendix D).
- Quantitive results suggesting the efficacy of HapticGen for vibrotactile design and qualitative insights about haptic designers' needs for a GenAI model from the three studies.

Finally, we make HapticGen and the datasets publicly available and open-source to support future research in this area. HapticGen is available at: *https://github.com/HapticGen/HapticGen*. The

two haptic datasets can be downloaded at: *https://github.com/HapticGen/hapticgen-dataset*.

2 Related Works

We review prior work in haptic design, generative models for visual and audio design conditioned on text descriptions, and the challenges and potential methods for constructing large-scale haptic datasets.

2.1 Haptic Signal Design Practices and Tools

In the last decade, haptics researchers have documented the difficulties of haptic design for experts [63] and novices [69]. These studies have revealed that programming haptic feedback using software development kits (SDKs) hinders creative ideation and rapid prototyping, which are essential to haptic design [45, 63]. They also highlighted the need for supporting design ideation in haptics, and provided requirements for the development of haptic design tools [64, 68]. Our work contributes to this literature with an account of the experience and needs of haptics experts and novices when designing with a generative text-to-haptics model.

Various graphical design tools have been proposed to facilitate haptic sensation design [14, 29, 68, 76, 77], including GUIs for vibrotactile devices [42, 55, 56, 64, 66, 82]. These tools enable designers to use direct manipulation to control the temporal patterns of one or more vibration actuators. According to a recent systematic review by Terenti and Vatavu [78], vibrotactile authoring tools may provide a library of previously designed vibrations [66, 71], allow designers to create and manipulate the waveform [56, 65, 66, 77, 78] or keyframes [64], and compose together multiple patterns [56, 65, 77, 78]. These tools facilitate creating and refining haptic signals by enabling rapid prototyping, but ideating haptic signals for a given application still requires design expertise, intuition and involves extensive trial and error [45, 63, 68].

Others have proposed audio-based design tools for haptics, leveraging the affinity between audio and vibrotactile signals. mHIVE uses temporal parameters of an audio signal, including Attack, Decay, Sustain, and Release (ADSR) in a GUI design tool to enable rapid prototyping of vibrations [65]. TECHTILE toolkit [52] allows designers to record audio (e.g., the sound of scratching a surface) and play it on one or more vibration actuators. Voodle [46] and Weirding Haptics [13] allow designers to record and convert their vocalizations into force or vibrotactile signals. Commercial tools such as Meta Haptics Studio [50] or bHaptics Studio [3] allow users to convert an audio file into vibrotactile signals for VR controllers and help further adjust the vibrations through a GUI editor. Drawing from previous works, we also leverage audio signals as a proxy for vibrotactile design but focus on using existing audio datasets to bootstrap a generative text-to-vibration model.

2.2 Enhancing Design Process with GenAI

Generative models conditioned on text descriptions have seen significant advancements and widespread use across various modalities, particularly in image creation. Text-to-image models, such as the DALL-E [59, 60, 72] and Stable Diffusion [16, 61] families, have revolutionized image generation by utilizing large-scale captioned image datasets. The success of these models has encouraged the exploration of generative design in audio and video modalities as well. A key challenge with these models often lies in the size and quality of the training datasets. For example, in developing DALL-E 3, researchers introduced a bespoke image recaptioning system and demonstrated that these synthetic captions improved prompt-following ability across several text-to-image models [72]. Aligned with previous findings, we synthesize additional captions for our dataset using an open-source large language model (LLM) to improve HapticGen's generation capability.

Recently, researchers have developed audio-captioned datasets and generative models for text-based audio design. AudioCaps, a strongly labeled subset of AudioSet [20], is a dataset of 46k audio snippets and their text descriptions collected through crowdsourcing [34]. Several other audio sources with descriptive tags also exist in the literature. WavCaps is a library of 400k audio snippets of 10 seconds with paired text captions harvested from various online sources and sound effect libraries, including AudioSet. The dataset uses the raw textual tags and descriptions from these sources, which are then filtered and transformed into homogeneous captions using ChatGPT [47]. Building on such datasets, researchers published AudioGen [39], an auto-regressive generative model that can generate short audio samples conditioned on text descriptions. The model is trained on 10 different text+audio datasets, including AudioCaps. MusicGen [6] is a text-to-music generative model that utilizes a similar transformer architecture as AudioGen but proposes an efficient training and signal codebook interleaving strategy to generate consistent music. Following this approach, the second version of AudioGen model (AudioGen v2) uses the same training strategy for generating audio snippets conditioned on text. To the best of our knowledge, no generative text-to-vibration model exists in the literature. A recent survey on the use of AI for extended realities (XR) highlighted the lack of prior work on applying AI to design haptic interactions in XR [25]. Yet, the advances in generative audio datasets and text-to-audio models have opened up new possibilities for haptics and inspired our work.

2.3 Efforts on Haptic Dataset Creation and Augmentation

One of the significant challenges in developing a generative haptic model is the scarcity and limited size of existing haptic datasets. While there are some open vibrotactile datasets available such as VibViz [71], these datasets typically contain only a few hundred signals compared to audio datasets such as AudioSet with over 2 million human-labeled audio signals [20] or vision datasets such as Microsoft COCO with over 300k images [44]. Additionally, many haptic datasets focus on textures and surfaces [8, 23], such as the LMT haptic texture database [75] or the Penn Haptic Texture Toolkit [9], which only represent a fraction of the vibrotactile signals that designers may wish to create. Notably, RecHap [80] explored augmenting a small dataset of hand-crafted mid-air ultrasound haptic signals to train a design recommendation system. Their approach relied on geometric transformations (scaling and rotation) to create variations, which may not provide enough diversity for a text-to-haptics generative model and also does not apply to our target vibrotactile devices with a single actuator.

Given the difficulty in hand-designing haptic signals, prior work has explored creating haptics from visual or audio content. In the visual domain, prior work has developed methods for video-to-haptics conversion. These approaches predominantly focus on capturing and generating motion-based or other highly spatial aspects [22, 30, 38, 41, 85]. Such approaches are useful for automating haptic creation for multiple (e.g., grids) of vibrotactile actuators [22, 30, 38] or for motion-based haptic devices (e.g., motion chairs) [41, 85]. We instead used audio-to-haptic conversion since this approach leverages the affinity between audio and vibration signals, has more research consensus, and better aligns with the capabilities of commodity haptic hardware, which typically features only a single vibrotactile actuator.

Although audio and vibrotactile signals are similar in digital representation, there are many perceptual differences to overcome. The perceptible range of audio is 20~20000 Hz while vibrotactile signals are perceptible around 0~1000 Hz, and the difference limen (just-noticeable difference) is also much finer for auditory pitch compared to tactile [7, 21]. Many methods have been explored for converting audio signals to vibrations [35, 43, 54, 86, 88]. Some rely on traditional signal processing techniques such as Okazaki et al. [54] while newer approaches leverage machine learning techniques to provide more powerful solutions. For example, Yun et al. developed a technique to classify video game sound effects and create multimodal vibration and impact feedback in real-time for first-person shooters and role-playing games [86]. Commercial solutions exist, such as those included in Interhaptics' Haptic Composer [27], bHaptics Studio [3], and Meta Haptics Studio [50], typically perform some form of amplitude mapping as well as detection of transients or impacts. Recognizing the diversity of existing methods, our training pipeline remains flexible and can easily be adapted to alternative or novel audio-to-haptic conversion approaches.

3 Formative Study: Haptic Desing Workshop with GenAI Tools

We started our process by conducting a design workshop with haptic researchers to explore whether and how state-of-the-art (SOTA) GenAI tools can be used for haptic design. We found several limitations and drawbacks of using existing GenAI tools, which we then tackled in HapticGen.

3.1 Participants and Procedure

A total of 9 participants voluntarily joined a 2-hour workshop held at a haptics conference. The participants included undergraduates (n=2, 22%), graduates (n=6, 67%), and a professor (n=1, 11%)). To ensure diversity, we recruited participants at various levels of haptic experience and paired junior researchers with senior ones. All participants, including the two undergraduates, had over one year of hands-on haptics experience in a research lab (average 2.2 years), in line with the criterion for haptics experts in the literature [87]. Most participants (n=5) specialized in vibration-based haptics, while others worked with force feedback (n=1) and thermal haptics (n=1). In terms of application domains, XR applications were the most common (n=7), followed by creating 4D contents (n=3) and games (n=3). We used Huggingface to provide access to publicly available GenAI tools including DALL-E [72], AudioGen [39], MusicGen [6], Make-an-Audio [26], Image-to-MusicGen [18]. After a brief introduction of the tools, participants collaborated in groups of 2~4 people to carry out haptic design tasks using GenAI tools. Specifically, they designed vibrotactile feedback based on 9 VR scenes spanning four categories: Physical Textures, Actions, Environmental Effects, and Emotional/Social interactions (e.g., petting a virtual dog). Besides GenAI tools, participants had access to basic multimedia tools for video, audio, and image editing (e.g., Quick time, Audacity, Adobe's creative tools), image or screen capture, and sound recording. Finally, we collected both quantitative and qualitative feedback from participants through a survey. The survey asked about the participants' design process, challenges encountered, and the utility of the GenAI tools. Participants were encouraged to think aloud during the process, and their responses were recorded and later analyzed by three organizers.

3.2 Results: Existing Challenges and Requirements for a Haptic Model

Participants faced several challenges in using GenAI models to create haptic effects. P2 and P6 tried the image-to-text or video-to-text models and found it difficult to create haptic effects based on the generated captions: *"It was difficult to find an appropriate model among video-to-text models. (P6)"* Most participants primarily used text-to-audio models (MusicGen, AudioGen), then converted the output into vibrations using Meta Haptics Studio. However, they also faced several challenges with these generative audio models while designing haptic experiences. We categorize their responses into three key takeaways regarding haptic designers' requirements for GenAI haptic models.

(1) Limitations of current GenAI models for haptic generation: Participants were often not satisfied with the text-to-audio model's output (n=4). P5 thought the issue was related to their limited prompting skills: *"We need to understand prompt engineering to get the results we want."*. Others also noted that the model did not follow their text prompts properly. P7 mentioned: *"With the current text-to-audio, the output often didn't match my intentions. "* P3 agreed: *"The GenAI didn't provide the answer we wanted, so we had to make adjustments."*. P4 felt the need for a set of "experiential keywords" to describe haptic effects and found the results of the generative audio models "poor". This challenge highlights the need for GenAI models that are specifically designed for haptics.

(2) Focusing on audio-to-haptic parameter tuning: Participants extensively revised signals and relied on trial-and-error with audio-to-haptic parameters to overcome limitations of the text-to-audio generative models (n=6). P1 said: *"I had to do a lot of manual work to achieve the desired output."*. P6 added: *"Modifying and improving prompts was challenging, and adjusting the generated audio to my liking was difficult."* P5 faced similar challenges while P2 observed, *"The effects of the parameters were somewhat unclear."* P7 and P9 also commented on the difficulty of understanding how audio conversion parameters impacted haptic effects. These difficulties were exacerbated since designers had limited knowledge of the audio-to-vibration conversion parameters in Meta Haptics Studio. This highlights that a GenAI haptic tool must incorporate the necessary pre- and post-processing steps to reduce cognitive load for designers.

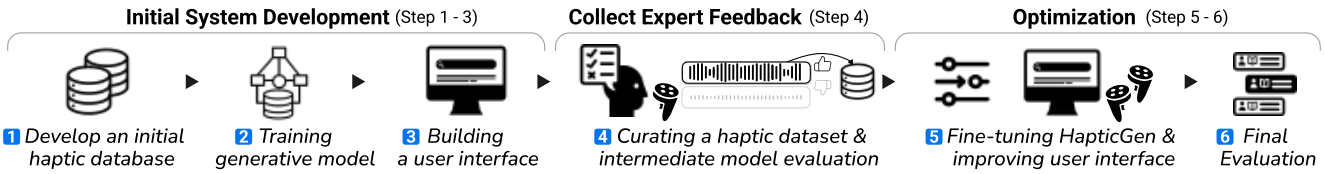


Figure 2: The overall process for designing and evaluating our system HapticGen.

(3) Overcoming obstacles in the design workflow: Participants faced challenges in testing the signals with the HuggingFace's basic interface (n=5). P1 mentioned, *"I encountered numerous limitations and complexities when using the [HuggingFace] UI."* P2 participant commented *"It seems there were many constraints and challenging aspects when it came to handling the user interface."*. P5 wanted to *"view the haptic design"* during generation and editing, and P5 and P7 needed easy playback functionality for testing. These feedbacks highlight significant usability issues, suggesting that the tool's interface needs improvement to allow for more accessible testing and preview of haptic effects. P9 also noted, *"The design tool lacks basic UI functionalities. For example, there's no way to play haptic effects without HMD"*. The inability to experience haptic feedback without additional hardware limits the tool's versatility and ease of use, potentially hindering the design process and user experience.

4 Overview of HapticGen Design and Evaluation Process

We iteratively developed HapticGen in the following steps (Figure 2):

Step 1 – Developing an initial haptic dataset from audio signals (Section 5.1): Given the lack of a large-scale text-vibration dataset, we bootstrapped our process by converting an existing text-audio dataset, namely WavCaps [47], into vibrations. Furthermore, we augmented the text labels to include tactile descriptions using an open-source LLM. This step resulted in an initial haptic dataset.

Step 2 – Training the generative model (Section 5.2): Next, we trained a SOTA autoregressive transformer model (MusicGen) on the converted haptic dataset from step 1, resulting in the first version of the text-to-vibration model or HapticGen.

Step 3 – Building a user interface (Section 5.3): To test the model, we iteratively developed a graphical UI for prompting the model, playing the vibrations on the VR controllers without needing to wear a VR HMD, and rating the match between the generated vibrations and the prompt using thumbs up/down buttons.

Step 4 – Curating a haptic dataset and intermediate model evaluation with haptics experts (Section 6): We ran a design study with 15 haptics experts using the HapticGen model and interface from the previous steps. In the study, the experts prompted the model through the UI, then they voted the generated vibrations with thumbs up or down. Finally, they answered questions about the utility of the model. This step resulted in an expert-voted haptic dataset and qualitative insights for improving the model and UI.

Step 5 – Fine-tuning HapticGen and improving user interface (Section 7): Next, we fine-tuned the HapticGen model using the expert-voted haptic dataset from the previous step. We also added pre- and post-processing steps and updated the UI to further improve the text-to-vibration design experience based on the experts' feedback.

Step 6 – Final evaluation (Section 8): We conducted an A/B testing study with 32 participants to evaluate the efficacy of HapticGen against a baseline model (MusicGen model plus audio-to-haptic conversion after inference).

This process resulted in HapticGen with three datasets; haptic-converted WavCaps – step 1, an expert-voted haptic dataset – step 4, a user-voted haptic dataset – step 6. A generative text-to-vibration model trained and fine-tuned on the first two datasets, and a user interface for prompting the model, playing the vibrations, and voting them. We used VR controllers from Meta Quest 3 [51] and pro [49] for all development and testing steps.

5 HapticGen: Dataset, Model, and Interface

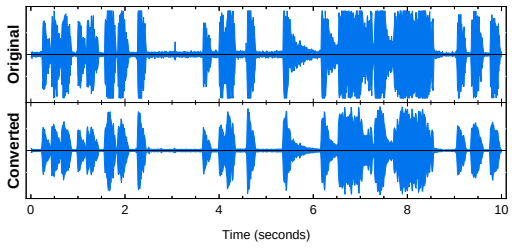
This section outlines the development of HapticGen, detailing the dataset creation, model architecture, and interface design. We provide an overview of the key processes: data preprocessing and conversion, model architecture and training, as well as interface design and functionality.

5.1 Initial Dataset: Haptic Converted WavCaps

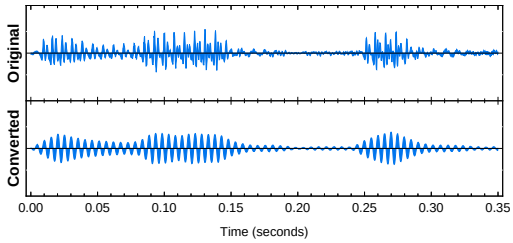
We built our initial haptic training dataset from the WavCaps [47] Audio Captioning Dataset (step 1 in Section 4) and later complemented it by collecting an expert-voted vibration dataset (step 4 in Section 4). WavCaps was chosen due to its size, ~400k audio clips with paired captions, and because it aggregates audio from multiple sources including FreeSound[1], BBC Sound Effects[2], SoundBible[3], and AudioSet[4]. To create a vibration dataset based on WavCaps, we followed three steps: filtering out speech, augmentation with haptic labels, and audio-to-haptic conversion.

5.1.1 Filtering Speech. Given that speech cannot be effectively replicated with vibrotactile feedback, training on samples containing speech-related content likely introduced noise, thereby increasing model perplexity without contributing to meaningful haptic generation. To address this, we filtered out any samples with labels containing terms such as "speech", "speak", "talk", "word", or

[1]https://freesound.org/
[2]https://sound-effects.bbcrewind.co.uk/
[3]https://soundbible.com/
[4]https://research.google.com/audioset/download_strong.html



(a) Full-length view of an original and converted training sample.



(b) Zoomed view of a 350ms segment of a signal highlighting detailed signal changes.

Figure 3: Two examples of original audio signals (top) and their converted haptic counterparts (bottom), using the process described in Section 5.1.3 – *Audio-to-Haptic Conversion*. The original audio signals are at 32 kHz, 16-bit resolution, and the converted vibrotactile signals are at 8 kHz, 8-bit resolution, with a controlled frequency.

"monologue". This filtering reduced the dataset size by approximately 15%, resulting in a final dataset size of ~335k samples (from ~400k samples originally).

5.1.2 Augmentation with Haptic Labels. To better align the audio captions used in training with the type of prompts a haptic designer might generate, we aimed to transform these captions to be more tactile-oriented. Given the very large size of the dataset, we leveraged an LLM-based approach for transforming the captions. Additionally, we found that supplying multiple caption variations per signal improved model performance. Our model was trained using the original caption and four additional tactile captions per signal. The additional tactile captions were created using Llama-3-8B-Instruct [15]. We devised four prompts to form the tactile caption variants. An example prompt asked, *"Write a single sentence that summarizes tactile feedback with the following attributes. Don't write anything about sound characteristics."* In this prompt, the "following attributes" referred to the original caption. See Appendix A for all four prompts and example captions.

5.1.3 Audio-to-Haptic Conversion. We decided to convert the audio signals to vibrations and train a model with the haptic-converted dataset rather than training on audio and converting signals to vibrations after inference. This approach allowed us to benefit from reduced inference latency (due to the lower sample rate, from 16 to 8 kHz), and enabled the use of expert-designed or manually-edited haptic signals in the training or fine-tuning process. We also anticipated improved performance with the reduced bit depth (8 bits for vibrations vs. 16 bits for audio) and lower overall complexity of the haptic signals compared to their audio equivalents.

In order to create vibration signals from the audio clips, we employed a technique similar to that demonstrated in Kim et al. [35], mapping the intensity and temporal characteristics of the audio signal onto a lower frequency sine wave to create a vibration (Figure 3). We did not use their proposed method for pitch matching as it may not apply well to our audio clip dataset, which contains longer, noisier audio clips with overlapping sound effects and complex frequency spectra. Instead, we relied mostly on the intensity and temporal characteristics of the signal to represent the haptic sensation due to the importance of these temporal rhythmic parameters in vibration perception [79]. We enabled some variation in frequency by mapping high and low intensities to a ±50 Hz range around the center frequency. Audio signals were broken up into 10-millisecond chunks, and the corresponding vibration signal was synthesized using a sinusoidal numerically controlled oscillator to support dynamic changes to frequency. The oscillator's base frequency is 220 Hz, which is around the highest human sensitivity to vibrations [21] and close to the middle of the voice coil motor's response capabilities in Meta Quest 3 and Pro VR controllers[5]. The final signals were sampled at 8 kHz and quantized to 8 bits, which aligns with the PCM output from the Meta Haptics Studio.

5.2 Model: Architecture and Training

In this section, we detail the architecture and training process of our model, outlined in Figure 4. Our approach leverages recent advancements in applying autoregressive transformer models to audio signals. We use this model architecture as a platform to integrate domain-specific modifications to optimize for haptic data generation and for further alignment with haptics expert preferences in Section 7.

We developed HapticGen by employing an autoregressive transformer-based decoder from the MusicGen architecture [6] and training it on our converted haptic dataset. This model works on quantized low frame rate tokens from an audio compression model such as EnCodec [12]. Specifically, we adapted the AudioGen v2 configuration, a reimplementation of AudioGen [39] that follows the architecture introduced in MusicGen. The code and solver for these models are available via the AudioCraft repository[6]. AudioGen v2 incorporates updates from MusicGen, including the use of a retrained EnCodec model and training on 10-second audio segments instead of 5 seconds. Although the model card notes that AudioGen v2 was trained without audio mixing augmentations, these augmentations are enabled by default in the current codebase, and we utilized them to train our model. The training samples are mixed by summing the waveforms at a random temporal offset, using a random signal-to-signal ratio in the range of [-5, 5], and concatenating the text captions. This augmentation enables the model to create complex compositions not explicitly present in the

[5]https://developer.oculus.com/documentation/unity/unity-haptics-apis/
[6]https://github.com/facebookresearch/audiocraft/

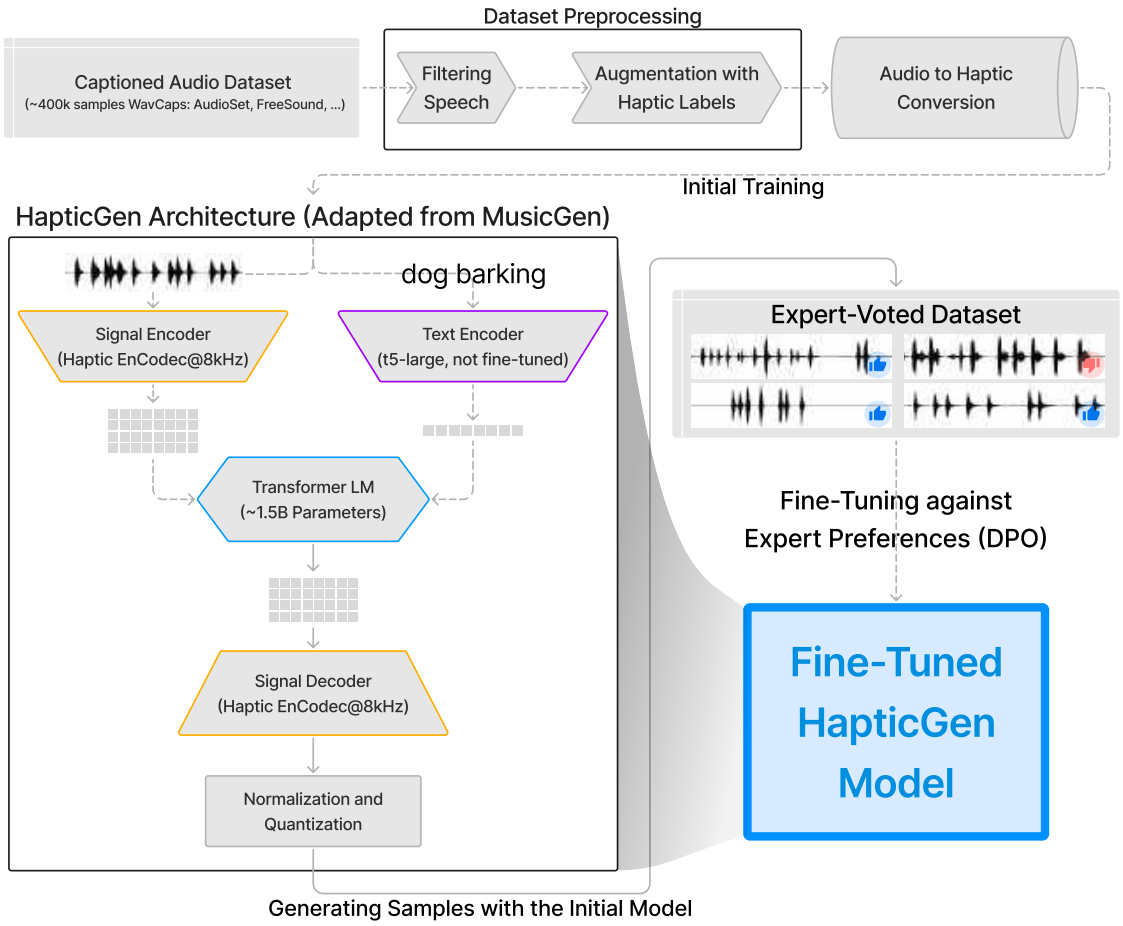


Figure 4: HapticGen Model Architecture and Training Overview - Due to the lack of a large, pre-existing haptic dataset, we began by filtering and applying automated transformations to an existing captioned audio dataset, generating corresponding vibrotactile signals and augmented haptic labels. We then used the resulting dataset to train the EnCodec tokenization model at a sample rate of 8 kHz. We trained the initial transformer language model (LM) using this EnCodec for signal tokenization and the haptic augmented labels for conditioning. Consistent with the MusicGen approach, the T5 text encoder was not fine-tuned for this task. During inference, the input prompt is processed by T5 and used as a conditioning input for the transformer LM. Output from the LM is decoded with EnCodec, then normalized and quantized to an 8 kHz 8-bit PCM signal for haptic playback. To further refine the model, we generated an expert-voted dataset of haptic signals and corresponding prompts using our initial model (Section 7.1). We employed this dataset to fine-tune the model using Direct Preference Optimization (DPO), leading to the final HapticGen model (Section 7.2).

training data. Besides mixing augmentations, the AudioGen v2 configuration uses the T5-large [58] text-to-text model for conditioning on a text prompt or label. For training the haptic model, we used the medium-sized transformer preset (~1.5B parameters) and lowered the sample rate from 16kHz to 8kHz to match our converted haptic dataset.

These training steps are outside the normal workflow for haptic designers. When a designer uses HapticGen for inference, their input prompt, for example "dog barking", is processed by T5 and used as a conditioning input for the pre-trained transformer language model. Output from the transformer is decoded using EnCodec, then normalized and quantized to an 8 kHz 8-bit PCM signal. This signal is then ready for the designer to play back on a haptic device, which we facilitated through a graphical interface, as discussed in the next section.

5.3 HapticGen Interface and Hardware

Based on the requirements from the formative study (Section 3), we developed a dedicated graphical user interface to facilitate haptic playback on Meta Quest Controllers and streamline the process of prompting and playing vibration signals from our model. As the Meta Quest controllers cannot be paired directly with a PC, and to avoid requiring designers to wear the head-mounted display (HMD), we developed a headless OpenXR application for the Meta Quest. This application forwards the controller inputs and haptic playback

features in real time to a browser-based application via WebSockets, enabling seamless use of the Meta Quest controllers without requiring any use of the HMD.

The interface is composed of four main components: the Generation, Results, Signal Browser, and Playback panes. The Generation pane enabled participants to prompt the initial model during the intermediate study (step 4 of Section 4) or prompt both the A and B models simultaneously during the final study (step 6 of Section 4). The system forwards prompts to cloud-hosted T4 GPU instances for model inference, and shows progress bars for the expected inference latency (typically 10-15 seconds). Once generated, the Results pane then displays the resulting signals for each model and provides the option to download them into a local folder for playback and voting. Then the user can select this folder in the Signal Browser pane to browse haptic signals, load them onto Quest Controllers for playback, and provide feedback by voting on the signals. The Playback pane includes a visualization of the haptic vibration currently loaded onto the Quest Controllers, accompanied by a real-time playback indicator. The footer displays the state of the WebSocket connection between the interface, web server, the Meta Quest device and controllers. For the user studies and data collection, the interface includes a participant ID input, enabling cloud storage of the generated signals and votes, facilitating easy retrieval of the dataset for model fine-tuning.

For example, a designer's workflow includes entering a prompt in the generation pane (e.g., "dog barking"). This prompt is then submitted to the cloud-hosted HapticGen model for inference, which responds with a list of possible generations. Designers can then save this set of signals locally, play each signal using the Quest controllers, then rate each signal with a thumbs up or down vote.

6 Curating a Haptic Dataset and Intermediate Model Evaluation with Haptics Experts

With our initial model, we held the design workshop to collect an expert-voted vibration dataset for model fine-tuning (Sec 7.1) and capture haptics experts' prompting practices and needs. Below, we report the workshop procedure and qualitative insights that helped improve our initial model and interface.

6.1 Methods

Participants. A total of 15 experts (5 females, mean age of 27) participated in the study. These included 4 industry experts working at an international haptics and VR company, and 11 participants from a haptics research group. They had worked with various haptic technologies, including vibrotactile (n=12), thermal (n=1), and force feedback (n=2). All participants had over one year of experience in designing haptics, similar to prior work [87].

Procedure. The study sessions were conducted at the company and a university research lab and took about 70 minutes on average. Participants used Quest Pro Controllers, HapticGen Model, and Interface to complete the study. They were asked to think aloud during the session.

Each session included a brief introduction (5 min), main design tasks (50 min), and survey & interview (15 min). After completing an informed consent, the experimenter briefly introduced the HapticGen interface and asked the participant to enter and play example prompts as a warmup task (e.g.,*"Lightly tapping on the mechanical keyboards as you are writing up a report.*"). During the main design tasks, participants chose one or more themes (Activity, Sports, Simulation, Emotions) and freely prompted HapticGen to generate signals. For each prompt, the system showed three vibration outputs with HapticGen. The participants tested these vibrations on Quest controllers and rated them with thumbs-up and down on the interface. In the end, participants completed the survey to describe their approach to designing haptic signals, suggest improvements for the model, and evaluate the overall experience. Industry experts (P1–P4) were also interviewed to provide insights into the haptic design and evaluation processes in their companies.

6.2 Results: Insights from the Experts

One author categorized all the answers and summarized the responses to reflect the participants' opinions. Participants identified areas for improvement and shared their thoughts while creating an average of 32 designs in an hour using the initial HapticGen model.

The four industry experts commented on their typical haptic design process for VR games. They noted that design was usually an individual process in their company, with collaborative evaluations occurring only when necessary. The primary assessment criteria were whether the design *"effectively reflects one's intention" and "fits well with the scene."* P1 explained: *"An individual designer first plays the target game scene and concisely notes the necessary elements before designing."* P3 added: *"designers often drew from previous haptic signals or their existing library and emphasized the need for solutions that can address abstract and ambiguous situations, such as social or emotional contexts, not covered in their current library."*

Participants found the model was sensitive to their input prompt. P2 noted: *"The more detailed the prompt, the closer it feels to what is desired."* The participants gradually added more specific situation descriptions or adjectives and adverbs to reflect their intents. Some participants found prompting challenging. For instance, P1 reported that *"It usually takes a longer time to make decisions about situations one has not experienced"* which indicates their difficulties in verbally expressing corresponding tactile representation. It was suggested that using an LLM to present multiple variations could be beneficial.

Participants had different criteria when evaluating HapticGen signals. Multiple participants mentioned that they compared the match between the vibration with what they had imagined for the situation (n=8) or the prompt (n=4). For instance, P6 said: *"[I downvoted] when the vibration is difficult to match with imagination."* and P1 noted: *"If some part of the signal give immersive experience, I rate it as good even."* Others mentioned factors such as feeling natural, good, and comfortable (P2–P4), immersive (P1, P5, P10), or realistic (P10, P12). Several participants also noted that they downvoted vibrations that had low intensity or were imperceptible (n=5). P4 pointed out, *"Overall intensity could be higher."* Relatedly, P2 noted that *"if the magnitude could be adjusted and amplified, it [HapticGen] would be highly applicable in various contexts."*, and P7, P8, P13 made similar comments. Based on these results, we improved the HapticGen model in several ways, including fine-tuning the model, facilitating user prompting by creating prompt variations with an LLM, and normalizing the signals from HapticGen. We provide further details in the next section.

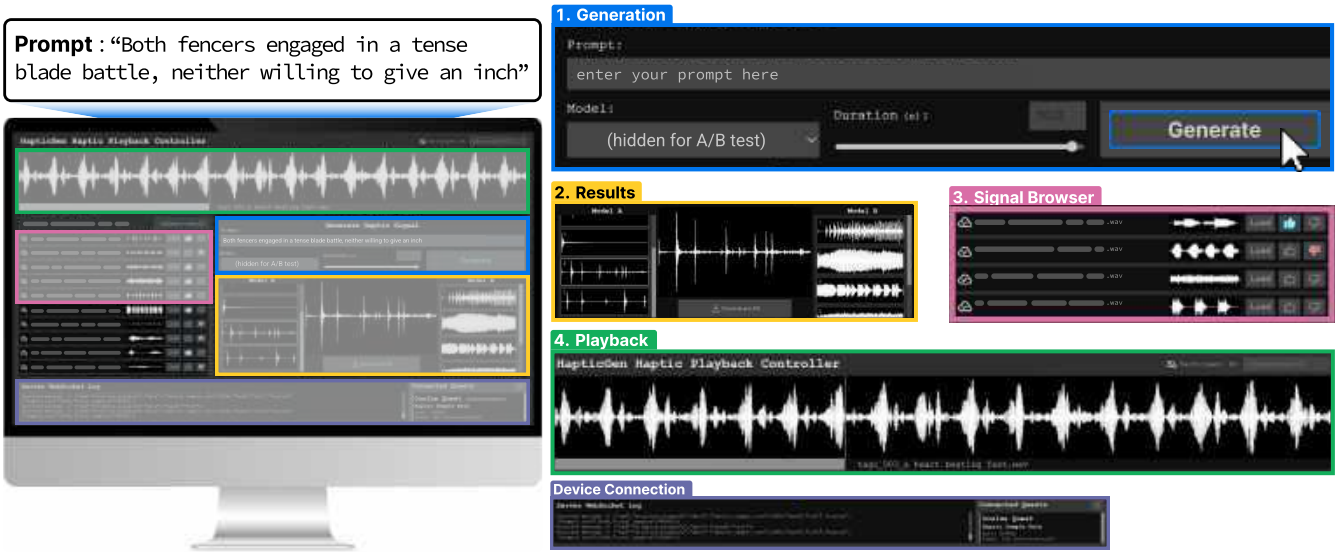


Figure 5: HapticGen Interface with the (1) Generation, (2) Results, (3) Signal Browser, and (4) Playback panes. This system consolidates the generation, playback, and evaluation of haptic signals into a single, seamless experience using the haptic actuators in Meta Quest controllers, eliminating the need for the HMD or multiple separate interfaces. The bottom of the UI displays information about the connected Quest and controllers.

Validation Dataset	Initial	SFT(NS)	**DPO**	IPO	ROBUST
Original Haptic Converted WavCaps (~33k)	0.00000	0.12636	0.00060	0.00061	0.00058
ExpertPref Mix (243 Voted + 1500 WavCaps)	0.00000	-0.59786	-0.42901	-0.42201	-0.42765

Table 1: Comparison of the change in validation cross-entropy (negative values indicate improved performance) across different fine-tuning methods on two datasets: the original Haptic Converted WavCaps validation split (~33k samples), and the ExpertPref Mix validation dataset, which contains 243 expert-voted samples combined with 1500 samples from the Haptic Converted WavCaps split. The 'Initial' model (before fine-tuning) is compared against standard Supervised Fine-Tuning with Negative Sampling (SFT(NS)), Direct Preference Optimization (DPO), Identity Preference Optimization (IPO), and Robust DPO (ROBUST).

7 Fine-Tuning HapticGen for Alignment with Expert Votes

We fine-tuned HapticGen on the expert-voted haptic dataset from the workshop to improve its quality of haptic signal generation. We also applied a set of pre- and post-processing steps to facilitate user prompting and reduce quiet signal generation based on the expert workshop.

7.1 Expert-Voted Haptic Dataset

As part of the above design workshop with haptics experts, we collected a dataset of vibration signals generated and voted by the expert haptic designers using our initial model. This dataset contains 1297 data points (prompt, signal, thumbs up/down vote) with 329 unique prompts. Specifically, for each prompt, the model generated multiple signals to rate (thumbs up or down). For fine-tuning, pairs of signals with the same prompt and contrasting votes (i.e., a vibration with thumbs up and another vibration with thumbs down) can be used to better guide the model optimizer. Our dataset contains 220 unique prompts with such direct signal preferences. Considering all possible combinations of paired signals, there were 1020 training samples with such paired preference. Of the prompts without paired preferences, 49 had only positive votes, and 60 had only negative votes.

7.2 Fine-Tuning Process and Results

To incorporate the expert-voted data, we integrated standard supervised fine-tuning (SFT) with negative sampling into the MusicGen solver. While SFT is effective in learning from labeled data, it has inherent limitations, such as susceptibility to catastrophic forgetting and overfitting. Therefore, in addition, we implemented Direct Preference Optimization (DPO) [57], which is a way to optimize for human preferences without requiring a separate reward model as in reinforcement learning. We also tested with Identity Preference Optimization (IPO) [2] and Robust DPO (ROBUST) [5] loss, but did not see significant differences or improvements in the model's validation metrics compared to standard DPO. DPO, IPO, and ROBUST were trained using 1020 paired preference samples while SFT was trained using all 1297 voted samples.

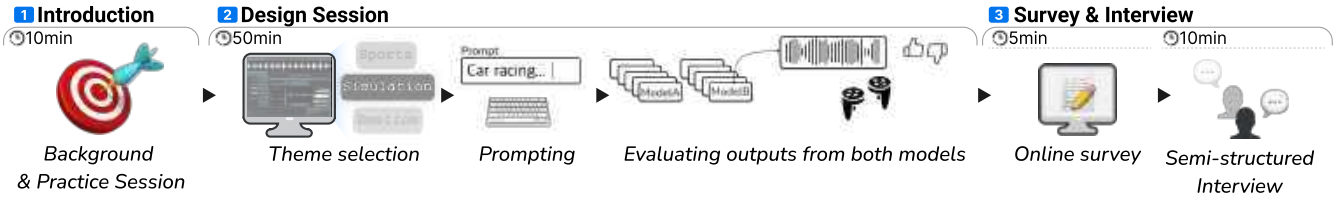


Figure 6: Study procedure of the final evaluation with 32 participants to evaluate user satisfaction with the fine-tuned HapticGen model and create a haptic dataset with over 3.2k vibrations.

The validation dataset for these models included a non-overlapping random selection of 243 expert-voted signals and 1500 signals from the original Haptic Converted WavCaps training validation split. We call this validation set ExpertPref Mix. We also validated the fine-tuned models on the original validation split to check for regression. We found the DPO models tended to diverge quickly on the expert-voted dataset, so we blended the positive sample cross-entropy loss with a factor of 0.01 for DPO, IPO, and ROBUST, which led to more stable results. In our final DPO, IPO, and ROBUST models, we used hyperparameters $\beta = 0.1, \tau = 0.1$, and for ROBUST we used $\varepsilon = 0.1$. Based on the results in Table 1, we determined that the DPO model offered a favorable balance between maintaining performance on the original Haptic Converted WavCaps validation dataset while demonstrating improvements on the ExpertPref Mix dataset. We used cross-entropy loss to compare the models since, in contrast to audio, the field of haptics does not have objective metrics for evaluating haptic signal quality. As the cross-entropy loss values were relatively close across the different methods (especially between DPO, IPO, and ROBUST), any of the fine-tuned models could potentially perform well in user evaluations. Still, we selected DPO out of the three preference optimization methods (DPO, IPO, ROBUST) since DPO showed the best improvement (i.e., a larger decrease in loss value) on the ExpertPref Mix. Although the gains on the ExpertPref Mix dataset were moderate compared to SFT, the DPO model demonstrated minimal regression on the original validation set, which was especially important given the relative size of the expert-voted dataset (1297 signals) compared to the original training set (~335k signals). This stability suggested that the DPO model was less likely, compared to the SFT model, to encounter issues such as overfitting, objective misalignment, bias amplification, or other undesirable behaviors during the final subjective user evaluation.

7.3 Facilitating User Prompting

In addition to fine-tuning the model, we adjusted HapticGen to facilitate text prompt creation based on feedback regarding prompting challenges and prompt sensitivity. We added a prompt preprocessing step to the HapticGen interface to create multiple variants of the input prompt, using an LLM, to further increase the variety of generated signals and assist when designers provide very brief prompts. These prompt variants are not visible to the designer but are used internally to enhance the diversity and variability of signals generated in each iteration. The variants are created using gpt-4o-mini-2024-07-18 [48] with the following prompt: *"Generate {n_variants} unique caption variants based on an input prompt for a generative model. Use clear and natural 3rd person language. Avoid creative flourishes and stick to straightforward captions. Avoid repetitive language and focus on creating a variety that covers the spectrum of possible generations."* The API call with this prompt requests structured output as a JSON string array. This prompt is haptic agnostic as it was also used for the baseline audio model in our final A/B evaluation.

7.4 Post-Processing Haptic Generations

Feedback from workshop participants highlighted that the haptic model sometimes produced very quiet or imperceptible vibrotactile signals on the Meta Quest controllers. To address this, we modified a post-processing component applied before signal quantization to prevent quiet or silent generations. By default, MusicGen compresses the decoded signal output before quantization to avoid clipping by scaling against the peak amplitude with 1dB headroom. To ensure perceptibility, we enabled signal normalization and implemented an amplification limit, setting the maximum normalization to -10dB, as a majority of signals below this threshold were voted negatively by haptics experts.

8 Final Evaluation of HapticGen

We ran an IRB-approved within-subject A/B testing study to compare the fine-tuned HapticGen model to a baseline generative audio model using a controlled study setting (e.g., user interface, hardware). We observed users' behaviors and collected their satisfaction with the model output using both quantitative and qualitative methods. Our primary focus was to evaluate the performance of the core text-to-vibration model itself, without integrating basic editing tools (e.g., temporal cropping, amplitude adjustment) commonly used in real-world workflows. To encourage a broad exploration of the model's capabilities, we provided examples of common XR themes rather than specific scenes that could limit creativity. This approach allowed us to emphasize model performance while also collecting a large and diverse dataset of rated signals per user.

8.1 Methods

8.1.1 Participants. A total of 32 participants (16 females, mean age of 26.1) took part in the study. Since the formative and intermediate studies revealed that haptic experiences are highly subjective, we included both expert (n=17, 53.1%) and novice designer groups (n=15, 46.9%). To ensure diversity, we recruited participants from various research backgrounds and majors, including human-computer interaction (n=12, 37.5%), haptics (n=5, 15.6%), industrial design (n=5, 15.6%), computer science (n=3, 9.4%) and others (n=7, 21.9%).

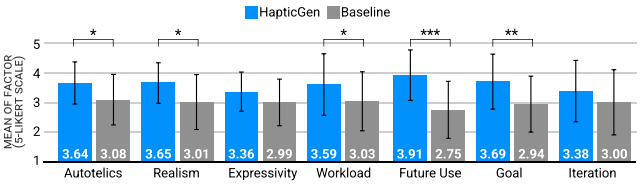
8.1.2 Apparatus and Setup. In our study, we compared two models: the text-to-vibration model (HapticGen) and a baseline text-to-audio model (AudioGen v2) followed by audio-to-vibration conversion using the same method as described in Section 5.1.3. The baseline presents the best existing GenAI approach available to haptic designers before HapticGen. Similar to HapticGen, the interface creates variations to the user prompt and presents five signals corresponding to these prompt variations for the baseline model. We did not apply the post-processing step described in Section 7.4 to the baseline as it required modifying the model's source code, which would make it not representative of the implementation currently available to designers. Moreover, this change was specifically developed to address the requirements of the haptic model. The baseline model was, on average, twice as loud as HapticGen, therefore additional normalization was unlikely to enhance its performance. This setup allowed us to conduct a head-to-head comparison between the performance of a tailored haptic model against a standard audio model in generating vibration feedback. Participants held the controller with their non-dominant hand and used the other hand to interact with the HapticGen interface (Fig 5).

8.1.3 Procedure. The study took 75 minutes per participant, starting with the introduction of HapticGen interface and the overall design process with examples such as *"The basketball player dribbles the ball"* and "*Trying to slice a tough steak"* for warmup tasks (Figure 6).

In the main design session, participants designed vibrations for five themes: Sport, Interaction, Emotion, Game, and Simulation. These themes were from intermediate evaluation with haptics experts (Section 6). We provided examples such as *"Trying to open a tightly shut pickle jar and then finally opening it with a loud pop!"*. After entering a prompt, they waited for 10~15 seconds to get 10 haptic feedback signals (5 each from models A and B). Models A and B were preset in the backend system as either baseline or HapticGen model. We counterbalanced the order of the models to minimize bias. Saved signals in the local folder were listed with prompts as file names for participants to review. They loaded and played the signals on Quest Controllers and voted with thumbs up and down. For each theme, participants created two prompts: 1) a free prompt, and 2) a modified version to refine their intentions.

After the tasks, we conducted a post-study questionnaire and a semi-structured interview to gather feedback on participants' prompting strategies, impressions of the model, and subjective experiences.

8.1.4 Data Collection and Analysis. In this study, we collected and analyzed both quantitative user ratings and qualitative comments. We obtained user ratings on Factors of Haptic Experience (HX) [1, 62] and the system usability. The ratings included Autotelics, Expressivity, and Realism from HX, Workload, and Future Use from [84], and Goal and Iteration from [11] (see Appendix C). A Shapiro-Wilk normality test showed that Autotelics, Realism, Expressivity, Workload, and Future Use followed a normal distribution ($p > 0.05$). Thus, we used paired t-tests to analyze these ratings. The ratings for Goal and Iteration did not follow a normal distribution, thus we used the Wilcoxon signed-rank test (a non-parametric test) instead. The analysis focused on various factors and questionnaire responses, with significant differences ($p < 0.05$) highlighted in Fig 7. For qualitative comments, the interviews were transcribed. Then, one author counted the words and extracted the main topics from each interview question. All four authors discussed and summarized them in writing.



(a) User ratings

HX & System Factor		Statistic	p		Effect size
Autotelics	Student's t	2.58	.015	Cohen's d	.456
Realism	Student's t	2.61	.014	Cohen's d	.462
Exoressivity	Student's t	1.99	.055	Cohen's d	.352
Workload	Student's t	2.09	.045	Cohen's d	.369
Future Use	Student's t	3.82	<.001	Cohen's d	.678
Goal	Wilcoxon W	2.46	.02	Rank biserial correlation	.435
Iteration	Wilcoxon W	1.51	.146	Rank biserial correlation	.266

(b) Results of statistical tests

Figure 7: Results of the final evaluation, showing (a) user ratings with significant differences marked with an asterisk, (b) results of paired t-tests (first five factors) and Wilcoxon signed-rank test (Goal, Iteration). Factors that show significant results at $p < .05$ level are bold-faced.

8.2 Quantitative Results

We performed statistical analysis and the results are shown in Figure 7. Participants perceived our system significantly better than the baseline across several quantitative metrics. Regarding HX factors, which measure satisfaction with the model's haptic output, participants largely preferred our system (Autotelics: $p = 0.015$, Realism: $p = 0.014$, and Expressivity: $p = 0.055$). They also found that our system supported more meaningful haptic design and reduced the workload for them (Workload: $p = 0.045$). Moreover, participants believed our model achieved the design goal more effectively than the baseline (Goal: $p = 0.02$). Notably, most participants expressed greater willingness to use our system over the baseline in the future (Future Use: $p < 0.001$). 75% of participants were satisfied with 2 or more signals out of HapticGen's 5 outputs. Of these, 28% were satisfied with 3 or more signals and 3% with 4 or more. In contrast, only about 53% of participants showed satisfaction with 2 or more signals for the baseline. The percentage satisfied with 3 or more signals from baseline was 19%, and no one was satisfied with 4 or more signals. This shows an improvement of over 40% satisfaction in the generated vibrations from HapticGen compared to the baseline.

8.3 Qualitative Results from Interviews

Prompting Strategies. Participants refined their haptic design prompts with three types of strategies: Scenario details, Expressive language, and Signal characteristics. Many participants enhanced the contextual richness of their prompts, as exemplified by P1's approach: *"I give more specific details about the effect that I want."* They put detailed scenario descriptions or object properties to give

hints to the model. For example, P31 initially wrote *Prompt : "writing letters on paper"* and later elaborated with situational details like *"a rough paper with pencil"*. This strategy allowed for more precise communication of desired outcomes. In terms of expressive language, participants significantly expanded their vocabulary or the way they expressed the target scene using adjectives and adverbs. P3 noted, *"[I] added more intense adverbs and scene description."* while P15 employed *"used multiple words and many adjectives to provide a richer description."* These linguistic enhancements aimed to convey more nuanced and vivid haptic experiences. Other participants demonstrated a growing complexity in tailoring their prompts to better convey desired temporal effects in vibrations. P6 explained that *"I modified the prompts to convey more dynamic vibration patterns or rhythmic timing"*.

Strengths of HapticGen. The participants' responses highlighted several key aspects of HaptiGen's performance. P19 observed, *"I noticed [for HapticGen] there was more variation in the vibrations. It felt more diverse and nuanced."* indicating the model's ability to produce a wide range of sensations. Multiple participants (n=9) emphasized the natural and realistic feel of the vibrations, with P2 stating, *"This model [HapticGen] often produced more natural-feeling vibrations. It wasn't as mechanical as the other [AudioGen]."* The model's capacity to inspire creativity was also a recurring theme (n=11), exemplified by P19's comment such as *"Experiencing the haptic feedback really sparked my creativity."* Furthermore, the model's ability to understand and reflect user intent was highly praised (n=14). P32 remarked, *"It consistently produced haptic results closer to what I intended. It's like it could translate my thoughts into sensations."* These responses collectively suggest that HapticGen significantly enhanced the design experience by providing dynamic, realistic, and inspiring tactile sensations that aligned with user intents. Participants were also impressed by the quality of vibrations generated by the AI model. P4 found the tool *"incredibly helpful"* in addressing the challenges of creating desired vibration patterns, noting that *"[HapticGen] generates irregular patterns and tension is impressive."* The ability to generate short impact vibrations was another notable aspect, with P16 observing that *"The tool excelled at designing momentary vibrations rather than repetitive ones. It captured the desired feeling well."* Lastly, the tool's capacity for facilitating creative haptic experiences was evident with P4 suggesting that *"the Interaction and Emotion themes could be particularly useful. They might be especially helpful in creating cinematic video experiences."* P31 was impressed by how well the tool reflected foot sensations, *"considering the focus is typically more on feet in real-life scenarios."* These responses collectively underscore HapticGen's potential to accelerate haptic design through AI integration, momentary vibrations, imagination stimulation, and facilitation of creative experiences.

Weaknesses of HapticGen. The participants also noted areas where HapticGen and Baseline did not meet their needs. Participants reported the hardware limitation regarding subtle vibrations that the controller could not render well. For instance, physical interactions such as shaking hands, combing one's hair, and makeup with a brush produced small and gentle feedback, which was difficult to feel on the controllers. In these cases, the Baseline was often preferred by participants who prioritized strong intensity (n=8 out of 32). Some participants were also less satisfied when both models did not grasp the necessary details of the target scenario. P18 expected the model to *"understand the meaning without detailed explanation"* and P29 *"didn't write the full context of the scenario, but only wrote the key interaction."*. This indicates that some users expect the model to understand the underlying meaning of the context and generate vibrations, which currently do not meet user satisfaction levels. For example, participants described scenarios involving a change in intensity or speed (e.g., a vibration for running then walking), but the generated vibration did not reflect a transition between running (strong, fast) and walking (weak, slow). Finally, seven participants mentioned that *"Both models did not work properly when I asked for repeated effects."*. These included prompts such as: *"...shake hands and wave up and down three times"* or *"During the first 1-2 seconds, I can [want to] strongly feel the sensation"*. We reflect on these limitations in section 9.

8.4 Haptic Dataset from Final Evaluation

User-voted haptic data was collected as part of the final A/B testing study (Figure 8). This dataset includes 3229 data points (prompt, haptic signal, thumbs up/down vote) from both models, with 326 unique prompts. Pairs of signals with the same prompt and contrasting votes can be especially useful for training (see Section 7). Out of the total 326 unique prompts, 315 contained such direct signal preferences, while 5 prompts received only positive votes and 6 received only negative votes. Considering all possible combinations of paired signals, the dataset yields 6729 paired preference training samples. We did not incorporate this dataset into HapticGen, however, the same model alignment process (Section 7) could be used to further improve haptic signal generation in future models. Both haptic datasets from the intermediate and final study are publicly available at this link: https://github.com/HapticGen/hapticgen-dataset.

9 Discussion

Here, we reflect on the paper's contributions, discuss its implications for the field of haptics, and outline our limitations and avenues for future work in generative haptics.

9.1 Reflection on Results and Contributions

Utility of HapticGen Model and Interface. Our results suggest the efficacy of HapticGen in supporting vibrotactile design. We iteratively designed HapticGen and its user interface through formative and design studies with expert designers and users. Based on findings from these studies, the model and interface facilitate rapid prototyping, signal evaluation, and data collection. In the final evaluation, participants most frequently highlighted HapticGen's ability to capture user intent from text prompts (43.75%, 14 out of 32 participants) and provide design inspiration (34.4%, 11 out of 32 participants) with nuanced and dynamic vibrations. These two aspects are complementary where the former refers to helping designers realize their intended concepts, and the latter is about providing unexpected but relevant signals that facilitate serendipity and ideation in design. These factors led to improved ratings for haptic experience factors, Workload, and Goal. We also observed a strong user inclination toward Future Use for HapticGen, with the largest improvement in ratings (from 2.75 to 3.91), which is

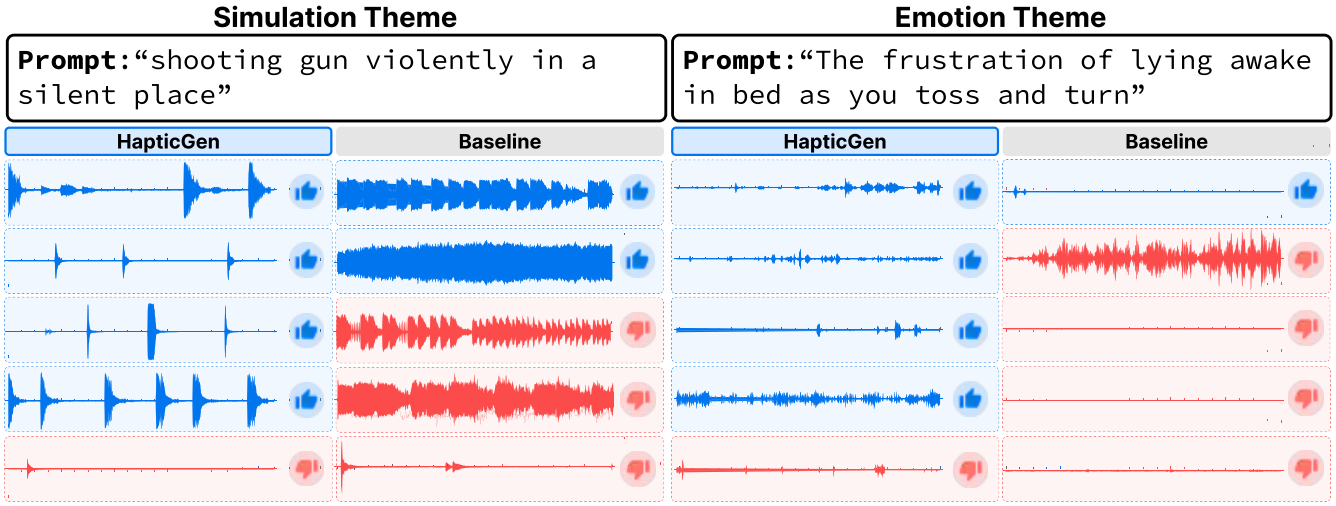


Figure 8: Example prompts provided by two participants, along with the corresponding vibrations generated by HapticGen and the Baseline model. Participants voted on the vibration outputs after experiencing the haptic feedback on Quest controllers.

perhaps the most important indicator of the model's utility. HapticGen's improved performance over the baseline approach can be attributed to three main factors. First, haptic signals are simpler in bit-depth, sample rate, and complexity compared to audio, making them more suitable for tokenization and prediction by the model. Second, we applied haptic-label augmentation prior to training, which improved the relevance of generated signals to user prompts. Finally, we fine-tuned the model using human-prompted, tested, and rated haptic signals, further aligning it with user preferences to generate high-quality vibrotactile signals. Still, opportunities for further refinement and alternative approaches exist, which we discuss in Section 9.3.

Challenges in Iterative Refinement. On the other hand, the Iteration ratings did not show a significant improvement over the Baseline. Although our interface helped support ideation and creating a diverse set of generations, we did not include any features for directly editing signals, which restricted refinement capabilities for this study to only prompt adjustments. Prompt modifications require a degree of intuition for prompt engineering, which could require longer-term experience with the model. Additionally, the iterative refinement process may have been impacted by the absence of functionality to condition new model outputs on previously generated signals, a technique explored in frameworks like TiGAN [89]. Such an approach could also mitigate instances where small prompt changes result in drastically different model outputs, a factor that may have further impeded iteration. This suggests that while the prompt-based approach offered flexibility for ideation, there may be value in incorporating more direct manipulation tools into the interface and model architecture to enhance the refinement process.

Captioned Haptic Datasets. Besides the model, the datasets collected through HapticGen represent the largest publicly available haptic datasets with text labels to date. This includes two human-voted datasets, which together contain over 4500 rated signals with 2088 receiving positive votes. This is one to two orders of magnitude larger than the number of signals in existing datasets like VibViz with 120 vibrations [71] or Feel Effects with 40 vibrations [28]. These large datasets can be valuable resources for developing predictive models of haptic experience or for training other haptic GenAI models. The distinction between the two datasets (expert-voted and user-voted) provides flexibility for future researchers, who can weigh them differently based on their specific goals or applications.

9.2 Implications of HapticGen for Haptic Design and Research

Streamlining Content Creation for Haptic Design. HapticGen addresses a key bottleneck in the haptic design process: the slow and labor-intensive creation of content. By enabling the rapid generation of vibrotactile signals from text inputs, HapticGen significantly accelerates the ideation process for haptic designers, allowing them to produce vibrotactile signals that can either be applied directly or further refined within traditional GUI-based design tools. Expert haptic designers observed that certain signal subsets generated by HapticGen were immediately applicable and often required only minor adjustments, such as trimming, before the final production. Designers may still desire additional control to refine a sensation using existing vibrotactile design tools such as those covered in Section 2.1. Such tools offer designers detailed control with features like direct waveform manipulation [56, 65, 66, 77, 78], especially in cases where changes may be hard to describe using natural language. In such tools, HapticGen can be integrated as an alternative to traditional design libraries, enabling faster navigation and exploration of the design space while also generating more diverse and nuanced designs. The HapticGen model can also be used together with existing audio-based design tools such as Weirding Haptics [13], enabling designers to use both vocalizations and speech to create vibration effects in VR. The ability of generative models like HapticGen to streamline the design process empowers

designers to explore a broader range of creative possibilities and produce high-quality, customized haptic content that moves beyond reliance on pre-existing libraries and lengthy iteration cycles.

Scaling Haptic Data Collection Practices with HapticGen. Generative haptic models also have broader implications for data collection in haptics research. This has traditionally been a challenge due to the designer skill and time required to create vibrations as well as the lack of a standard vocabulary for labeling vibrotactile signals. In the final evaluation, we were able to collect 3229 user-voted haptic signals within a relatively short time frame (averaging over 100 vibrations and votes per hour), significantly outpacing previous studies where users typically describe less than 10 to 15 signals per session [10, 53, 70]. This time-cost efficiency demonstrates that HapticGen could facilitate the creation of large human-labeled haptic datasets quickly. By targetting standardized commodity hardware, HapticGen also offers potential for scaling through crowdsourcing approaches [40, 67] to data generation and collection.

Haptic Hardware Extensibility. Beyond the immediate context of VR controllers, HapticGen's core signal generation methods are generalizable to other haptic modalities. While our work used voice coil actuators in Meta Quest controllers, the generated signals primarily convey information through variations in intensity and rhythm, making them compatible with other haptic devices such as linear resonant actuators (LRA) and eccentric rotating mass (ERM) motors. Additionally, the HapticGen training pipeline can easily be adapted to alternative audio-, video-, or other modality-to-haptic methods that may better suit other haptic hardware (e.g., force feedback or thermal) [4, 30, 35, 88]. For example, recent work on haptic force estimation from video [4] can be applied to existing kinetic or human-action video datasets [32, 74] to create large initial datasets for force-feedback technology where none currently exist. Following our approach, the time-series force data can be learned by a tokenization model and captions describing the physical interactions in the videos could be augmented using pre-trained LLMs. This dataset can help train an initial generative text-to-force feedback model, which one can further fine-tune using human-in-the-loop approaches as in HapticGen.

Applications of HapticGen Users across all expertise levels can create and customize haptic experiences effectively with HapticGen. Through its user-centric approach and intuitive prompt-based interface, the system lowers traditional technical barriers that have historically limited haptic design to specialists. For instance, automotive designers can rapidly prototype and test different haptic patterns for their specialized applications. With HapticGen, end users will have the flexibility to customize haptic notifications in both 2D mobile interfaces and 3D VR environments according to their preferences. XR developers can also seamlessly integrate haptic effects into their immersive experiences using platforms like Unity.

9.3 Limitations and Future Work

Audio-to-Haptic Conversion Methods. This work relied on an audio-to-haptic conversion method focused on signal intensity and rhythm characteristics, as these are the most essential in vibration perception [79, 81] and are also relatively straightforward to map. However, as noted in Section 2.3, a wide range of audio-to-haptic methods has been explored. Incorporating more advanced methods, particularly those that support perceptual mapping of frequency, could further improve the diversity of outputs from this model. Our dataset preprocessing and training pipeline fully supports alternative or novel conversion methods, and will just require retraining the EnCodec signal tokenizer to capture new signal characteristics, followed by retraining the transformer model to accommodate the updated codebooks and tokens.

Model Fine-Tuning Methods. For fine tuning, we utilized standard supervised-fine tuning with negative sampling as well as Direct Preference Optimization (DPO) [57] and the related Identity Preference Optimization (IPO) [2] and Robust DPO (ROBUST) [5] loss functions. These fine-tuning methods rely on paired preference samples, where at least two samples using the same text condition (prompt) had opposite votes. As such, we could not use expert-voted data where all samples for a prompt had only positive or negative votes with these approaches. Very recent methods for fine-tuning such as Kahneman-Tversky Optimization (KTO) [17] or Binary Classifier Optimization (BCO) [31], which implicitly minimizes DPO loss, could unlock further gains for the model using a similar data collection method, especially as they do not necessarily require pairwise data.

Limited Prompt Control. As participants noted, the model cannot consistently handle prompt requests that specify the quantity or timing of events in the signal (e.g., two gunshots vs. five gunshots or two seconds of ticking followed by an explosion). This likely arises because the original audio captions often lack such detailed temporal information. In addition, when the timing did not align with the imagined target audiovisual material and haptic feedback, users should manually edit or play the desired part to match the timing. Larger datasets could mitigate this issue, but future work might explore more advanced "recaptioning" approaches to better label the source audio dataset. Such approaches could leverage pre-trained audio classification models [83] to generate more detailed captions considering temporal characteristics.

Objective Metrics for Evaluating Haptics. In many fields, deep learning models are evaluated using objective metrics and benchmarks, such as the Massive Multi-Task Language Understanding (MMLU) benchmark for language models [24] or the Fréchet Audio Distance (FAD) metric for audio enhancement [33]. Unfortunately, vibrotactile signals lack comparable metrics, forcing us to rely on model loss metrics like cross-entropy to compare performance. Since collecting large-scale subjective ratings for haptics is difficult, especially given the subtle differences between model improvements, the development of objective metrics is critical. Any such metrics would need to be validated against human perception to ensure their effectiveness in evaluating haptic signals.

10 Conclusion

In this paper, we introduced HapticGen, the first generative model for creating vibrotactile haptic signals from text inputs. Our evaluation demonstrated that HapticGen enhances the ideation process, lowers the barrier to entry for haptic design, and improves perceived haptic experience, system usability and future adoption compared

to a baseline model. By leveraging a large audio dataset and collecting haptic expert preferences, HapticGen effectively addresses the challenge of limited haptic data, enabling designers to create expressive and perceptually rich haptic signals. Looking ahead, the open-source release of the HapticGen model and the accompanying datasets with fine-tuning capabilities provides a valuable resource for future research, with potential applications in expanding haptic interactions and further refining GenAI-driven design tools.

Acknowledgments

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Global Research Support Program in the Digital Field (RS-2024-00419561) supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP), and by the National Research Council of Science & Technology (NST) grant from the Korea government (MSIT) (No. CRC21014). It was also supported by VILLUM FONDEN (VIL50296) and the National Science Foundation (#2339707).

References

[1] Ahmed Anwar, Tianzheng Shi, and Oliver Schneider. 2023. Factors of haptic experience across multiple haptic modalities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–12.

[2] Mohammad Gheshlaghi Azar, Mark Rowland, Bilal Piot, Daniel Guo, Daniele Calandriello, Michal Valko, and Rémi Munos. 2023. A General Theoretical Paradigm to Understand Learning from Human Preferences. arXiv:2310.12036 [cs.AI] https://arxiv.org/abs/2310.12036

[3] bHaptics. 2021. bHaptics Studio. https://www.bhaptics.com/software/studio/

[4] Xiaoming Chen, Zeke Zexi Hu, Guangxin Zhao, Haisheng Li, Vera Chung, and Aaron Quigley. 2024. Video2Haptics: Converting Video Motion to Dynamic Haptic Feedback with Bio-Inspired Event Processing . *IEEE Transactions on Visualization & Computer Graphics* 30, 12 (Dec. 2024), 7717–7735. doi:10.1109/TVCG.2024.3360468

[5] Sayak Ray Chowdhury, Anush Kini, and Nagarajan Natarajan. 2024. Provably Robust DPO: Aligning Language Models with Noisy Feedback. https://openreview.net/forum?id=FDmiBg8aH1

[6] Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. 2023. Simple and Controllable Music Generation. In *Thirty-seventh Conference on Neural Information Processing Systems*. https://openreview.net/forum?id=jtiQ26sCJi

[7] Stanley Coren, Lawrence M Ward, and James T Enns. 2004. *Sensation and perception*. John Wiley & Sons Hoboken, NJ.

[8] Heather Culbertson, Juan José López Delgado, and Katherine J Kuchenbecker. 2014. One hundred data-driven haptic texture models and open-source methods for rendering on 3D objects. In *2014 IEEE haptics symposium (HAPTICS)*. IEEE, 319–325.

[9] Heather Culbertson, Juliette Unwin, and Katherine J Kuchenbecker. 2014. Modeling and rendering realistic textures from unconstrained tool-surface interactions. *IEEE transactions on haptics* 7, 3 (2014), 381–393.

[10] Tor Salve Dalsgaard, Joanna Bergström, Marianna Obrist, and Kasper Hornbæk. 2022. A user-derived mapping for mid-air haptic experiences. *International Journal of Human Computer Studies* 168 (12 2022). doi:10.1016/j.ijhcs.2022.102920

[11] 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 CHI Conference on Human Factors in Computing Systems*. 1–22.

[12] Alexandre Défossez, Jade Copet, Gabriel Synnaeve, and Yossi Adi. 2023. High Fidelity Neural Audio Compression. https://openreview.net/forum?id=ivCd8z8zR2 Featured Certification, Reproducibility Certification.

[13] Donald Degraen, Bruno Fruchard, Frederik Smolders, Emmanouil Potetsianakis, Seref Güngör, Antonio Krüger, and Jürgen Steimle. 2021. Weirding Haptics: In-Situ Prototyping of Vibrotactile Feedback in Virtual Reality through Vocalization. In *The ACM Symposium on User Interface Software and Technology*. 936–953.

[14] Alexandra Delazio, Ken Nakagaki, Roberta L Klatzky, Scott E Hudson, Jill Fain Lehman, and Alanson P Sample. 2018. Force jacket: Pneumatically-actuated jacket for embodied haptic experiences. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, and et al. 2024. The Llama 3 Herd of Models. arXiv:2407.21783 [cs.AI] https://arxiv.org/abs/2407.21783

[16] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. 2024. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. arXiv:2403.03206 [cs.CV] https://arxiv.org/abs/2403.03206

[17] Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. 2024. KTO: Model Alignment as Prospect Theoretic Optimization. arXiv:2402.01306 [cs.LG] https://arxiv.org/abs/2402.01306

[18] Sylvain Filoni. 2023. Image to MusicGen - a Hugging Face Space by fffiloni. https://huggingface.co/spaces/fffiloni/Image-to-MusicGen

[19] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for datasets. *Commun. ACM* 64, 12 (2021), 86–92.

[20] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *Proc. IEEE ICASSP 2017*. New Orleans, LA.

[21] A. Gescheider, S.J. Bolanowski, and K.R. Hardick. 2001. The frequency selectivity of information-processing channels in the tactile sensory system. *Somatosensory & motor research* 18, 3 (2001), 191–201.

[22] Daniel Gongora, Hikaru Nagano, Masashi Konyo, and Satoshi Tadokoro. 2017. Vibrotactile rendering of camera motion for bimanual experience of first-person view videos. In *2017 IEEE World Haptics Conference (WHC)*. 454–459. doi:10.1109/WHC.2017.7989944

[23] Waseem Hassan, Arsen Abdulali, Muhammad Abdullah, Sang Chul Ahn, and Seokhee Jeon. 2018. Towards Universal Haptic Library: Library-Based Haptic Texture Assignment Using Image Texture and Perceptual Space. *IEEE Transactions on Haptics* 11, 2 (2018), 291–303. doi:10.1109/TOH.2017.2782279

[24] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring Massive Multitask Language Understanding. https://openreview.net/forum?id=d7KBjmI3GmQ

[25] Teresa Hirzle, Florian Müller, Fiona Draxler, Martin Schmitz, Pascal Knierim, and Kasper Hornbæk. 2023. When XR and AI Meet - A Scoping Review on Extended Reality and Artificial Intelligence. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 730, 45 pages. doi:10.1145/3544548.3581072

[26] Rongjie Huang, Jiawei Huang, Dongchao Yang, Yi Ren, Luping Liu, Mingze Li, Zhenhui Ye, Jinglin Liu, Xiang Yin, and Zhou Zhao. 2023. Make-An-Audio: Text-To-Audio Generation with Prompt-Enhanced Diffusion Models. arXiv:2301.12661 [cs.SD] https://arxiv.org/abs/2301.12661

[27] Interhaptics. 2022. Audio to Haptics - How Interhaptics Works? https://www.interhaptics.com/tech/how-interhaptics-works/

[28] Ali Israr, Siyan Zhao, Kaitlyn Schwalje, Roberta Klatzky, and Jill Lehman. 2014. Feel Effects: Enriching Storytelling With Haptic Feedback. *ACM Transactions on Applied Perception (TAP)* 11, 3 (2014), 1–17.

[29] 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*. 1–15.

[30] Kyungeun Jung, Sangpil Kim, Seungjae Oh, and Sang Ho Yoon. 2024. HapMotion: motion-to-tactile framework with wearable haptic devices for immersive VR performance experience. *Virtual Reality* 28, 1 (2024), 13.

[31] Seungjae Jung, Gunsoo Han, Daniel Wontae Nam, and Kyoung-Woon On. 2024. Binary Classifier Optimization for Large Language Model Alignment. arXiv:2404.04656 [cs.LG] https://arxiv.org/abs/2404.04656

[32] Will Kay, Joao Carreira, Karen Simonyan, Brian Zhang, Chloe Hillier, Sudheendra Vijayanarasimhan, Fabio Viola, Tim Green, Trevor Back, Paul Natsev, Mustafa Suleyman, and Andrew Zisserman. 2017. The Kinetics Human Action Video Dataset. arXiv:1705.06950 [cs.CV] https://arxiv.org/abs/1705.06950

[33] Kevin Kilgour, Mauricio Zuluaga, Dominik Roblek, and Matthew Sharifi. 2019. Fréchet Audio Distance: A Metric for Evaluating Music Enhancement Algorithms. arXiv:1812.08466 [eess.AS] https://arxiv.org/abs/1812.08466

[34] Chris Dongjoo Kim, Byeongchang Kim, Hyunmin Lee, and Gunhee Kim. 2019. Audiocaps: Generating captions for audios in the wild. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 119–132.

[35] 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 (2024), 2–7. doi:10.1109/TOH.2023.3338224

[36] Erin Kim and Oliver Schneider. 2020. Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX. *Conference on Human Factors in Computing Systems - Proceedings*. doi:10.1145/3313831.3376280

[37] 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*. 1–13.

[38] Myongchan Kim, Sungkil Lee, and Seungmoon Choi. 2014. Saliency-Driven Real-Time Video-to-Tactile Translation. *IEEE Transactions on Haptics* 7, 3 (2014),

394–404. doi:10.1109/TOH.2013.58
[39] 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. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=CYK7RfcOzQ4
[40] Dongjae Kwon, Ramzi Abou Chahine, Chungman Lim, Hasti Seifi, and Gunhyuk Park. 2023. Can we crowdsource Tacton similarity perception and metaphor ratings? *Conference on Human Factors in Computing Systems - Proceedings*. doi:10.1145/3544548.3581120
[41] Jaebong Lee, Bohyung Han, and Seungmoon Choi. 2016. Motion Effects Synthesis for 4D Films. *IEEE Transactions on Visualization and Computer Graphics* 22, 10 (2016), 2300–2314. doi:10.1109/TVCG.2015.2507591
[42] Jaebong Lee, Jonghyun Ryu, and Seungmoon Choi. 2009. Vibrotactile Score: A Score Metaphor for Designing Vibrotactile Patterns. In *IEEE World Haptics Conference (WHC)*. IEEE, 302–307.
[43] Yaxuan Li, Yongjae Yoo, Antoine Weill-Duflos, and Jeremy Cooperstock. 2021. Towards Context-aware Automatic Haptic Effect Generation for Home Theatre Environments. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 13, 11 pages. doi:10.1145/3489849.3489887
[44] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 740–755.
[45] Karon E MacLean, Oliver S Schneider, and Hasti Seifi. 2017. Multisensory haptic interactions: understanding the sense and designing for it. In *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations-Volume 1*. 97–142.
[46] David Marino, Paul Bucci, Oliver S Schneider, and Karon E MacLean. 2017. Voodle: Vocal doodling to sketch affective robot motion. In *Proceedings of the ACM Conference on Designing Interactive Systems (DIS)*. 753–765.
[47] Xinhao Mei, Chutong Meng, Haohe Liu, Qiuqiang Kong, Tom Ko, Chengqi Zhao, Mark D. Plumbley, Yuexian Zou, and Wenwu Wang. 2024. WavCaps: A ChatGPT-Assisted Weakly-Labelled Audio Captioning Dataset for Audio-Language Multimodal Research. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2024), 1–15.
[48] Jacob Menick, Kevin Lu, Shengjia Zhao, and et al. 2024. GPT-4o mini: advancing cost-efficient intelligence. https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/
[49] Meta. 2022. Meta Quest Pro: Premium Mixed Reality. https://www.meta.com/quest/quest-pro/
[50] Meta. 2023. Haptics Studio | Oculus Developers. https://developer.oculus.com/resources/haptics-studio/
[51] Meta. 2023. Meta Quest 3: New Mixed Reality VR Headset. https://www.meta.com/quest/quest-3/
[52] Kouta Minamizawa, Yasuaki Kakehi, Masashi Nakatani, Soichiro Mihara, and Susumu Tachi. 2012. TECHTILE Toolkit: A prototyping tool for design and education of haptic media. In *VRIC '12*. ACM Press, New York, USA, 2. doi:10.1145/2331714.2331745
[53] Marianna Obrist, Sue Ann Seah, and Sriram Subramanian. 2013. Talking about tactile experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1659–1668.
[54] Ryuta Okazaki, Hidenori Kuribayashi, and Hiroyuki Kajimoto. 2015. *The Effect of Frequency Shifting on Audio–Tactile Conversion for Enriching Musical Experience*. Springer Japan, Tokyo, 45–51. doi:10.1007/978-4-431-55690-9_9
[55] Sabrina Panëels, Margarita Anastassova, and Lucie Brunet. 2013. Tactiped: Easy Prototyping of Tactile Patterns. In *IFIP Conference on Human-Computer Interaction*. Springer, 228–245.
[56] E. Pezent, B. Cambio, and M. K. O'Malley. 2020. Syntacts: Open-Source Software and Hardware for Audio-Controlled Haptics. *IEEE Transactions on Haptics (ToH)* (2020).
[57] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. https://openreview.net/forum?id=HPuSIXJaa9
[58] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 1, Article 140 (jan 2020), 67 pages.
[59] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125 [cs.CV] https://arxiv.org/abs/2204.06125
[60] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. arXiv:2102.12092 [cs.CV] https://arxiv.org/abs/2102.12092
[61] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV] https://arxiv.org/abs/2112.10752
[62] Suji Sathiyamurthy, Melody Lui, Erin Kim, and Oliver Schneider. 2021. Measuring Haptic Experience: Elaborating the HX model with scale development. In *2021 IEEE World Haptics Conference (WHC)*. IEEE, 979–984.
[63] 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 (2017), 5–21.
[64] Oliver S Schneider, Ali Israr, and Karon E MacLean. 2015. Tactile Animation by Direct Manipulation of Grid Displays. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 21–30.
[65] Oliver S Schneider and Karon E MacLean. 2014. Improvising design with a haptic instrument. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 327–332.
[66] Oliver S Schneider and Karon E MacLean. 2016. Studying Design Process and Example Use With Macaron, a Web-Based Vibrotactile Effect Editor. In *IEEE Haptics Symposium (HAPTICS)*. IEEE, 52–58.
[67] Oliver S. Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E. MacLean. 2016. HapTurk: Crowdsourcing Affective Ratings of Vibrotactile Icons. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3248–3260. doi:10.1145/2858036.2858279
[68] Hasti Seifi, Sean Chew, Antony James Nascè, William Edward Lowther, William Frier, and Kasper Hornbæk. 2023. Feellustrator: A Design Tool for Ultrasound Mid-Air Haptics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
[69] 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 (ToH)* 13, 4 (2020), 791–805.
[70] Hasti Seifi and Karon E MacLean. 2017. Exploiting haptic facets: Users' sensemaking schemas as a path to design and personalization of experience. *International Journal of Human-Computer Studies* 107 (2017), 38–61.
[71] Hasti Seifi, Kailun Zhang, and Karon E. MacLean. 2015. VibViz: Organizing, visualizing and navigating vibration libraries. In *2015 IEEE World Haptics Conference (WHC)*. 254–259. doi:10.1109/WHC.2015.7177722
[72] Zhan Shi, Xu Zhou, Xipeng Qiu, and Xiaodan Zhu. 2020. Improving Image Captioning with Better Use of Captions. arXiv:2006.11807 [cs.CV] https://arxiv.org/abs/2006.11807
[73] Tanay Singhal and Oliver Schneider. 2021. Juicy Haptic Design: Vibrotactile Embellishments Can Improve Player Experience in Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 126, 11 pages. doi:10.1145/3411764.3445463
[74] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. 2012. UCF101: A Dataset of 101 Human Actions Classes From Videos in The Wild. arXiv:1212.0402 [cs.CV] https://arxiv.org/abs/1212.0402
[75] Matti Strese, Jun-Yong Lee, Clemens Schuwerk, Qingfu Han, Hyoung-Gook Kim, and Eckehard Steinbach. 2014. A Haptic Texture Database for Tool-mediated Texture Recognition and Classification. In *Proc. of IEEE Int. Symposium on Haptic Audio-Visual Environments and Games (HAVE)*.
[76] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–28.
[77] Colin Swindells, Evgeny Maksakov, Karon E MacLean, and Victor Chung. 2006. The role of prototyping tools for haptic behavior design. In *IEEE Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, 161–168.
[78] Mihail Terenti and Radu-Daniel Vatavu. 2023. VIREO: Web-based Graphical Authoring of Vibrotactile Feedback for Interactions with Mobile and Wearable Devices. *International Journal of Human–Computer Interaction* 39, 20 (2023), 4162–4180. doi:10.1080/10447318.2022.2109584 arXiv:https://doi.org/10.1080/10447318.2022.2109584
[79] David Ternes and Karon E. MacLean. 2008. Designing Large Sets of Haptic Icons with Rhythm. In *Haptics: Perception, Devices and Scenarios*, Manuel Ferre (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–208.
[80] Karthikan Theivendran, Andy Wu, William Frier, and Oliver Schneider. 2023. RecHap: An Interactive Recommender System For Navigating a Large Number of Mid-Air Haptic Designs. *IEEE Transactions on Haptics* (2023), 1–12. doi:10.1109/TOH.2023.3276812
[81] Jan BF van Erp, Michiel MA Spapé, et al. 2003. Distilling the underlying dimensions of tactile melodies. In *Proceedings of Eurohaptics*, Vol. 2003. 111–120.
[82] Dennis Wittchen, Bruno Fruchard, Paul Strohmeier, and Georg Freitag. 2021. TactJam: a collaborative playground for composing spatial tactons. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 1–4.
[83] Liang Xu, Lizhong Wang, Sijun Bi, Hanyue Liu, and Jing Wang. 2023. Semi-supervised sound event detection with pre-trained model. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[84] Yilin Ye, Qian Zhu, Shishi Xiao, Kang Zhang, and Wei Zeng. 2024. The Contemporary Art of Image Search: Iterative User Intent Expansion via Vision-Language Model. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 180 (apr 2024), 31 pages. doi:10.1145/3641019

[85] Gyeore Yun, Hyoseung Lee, Sangyoon Han, and Seungmoon Choi. 2021. Improving Viewing Experiences of First-Person Shooter Gameplays with Automatically-Generated Motion Effects. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 320, 14 pages. doi:10.1145/3411764.3445358

[86] Gyeore Yun, Minjae Mun, Jungeun Lee, Dong-Geun Kim, Hong Z Tan, and Seungmoon Choi. 2023. Generating Real-Time, Selective, and Multimodal Haptic Effects from Sound for Gaming Experience Enhancement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 315, 17 pages. doi:10.1145/3544548.3580787

[87] Gyeore Yun, Minjae Mun, Jungeun Lee, Dong-Geun Kim, Hong Z Tan, and Seungmoon Choi. 2023. Generating real-time, selective, and multimodal haptic effects from sound for gaming experience enhancement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[88] Yan Zhan, Xiaoying Sun, Qinglong Wang, and Weizhi Nai. 2024. Method for Audio-to-Tactile Cross-Modality Generation Based on Residual U-Net. *IEEE Transactions on Instrumentation and Measurement* 73 (2024), 1–14. doi:10.1109/TIM.2023.3336453

[89] Yufan Zhou, Ruiyi Zhang, Jiuxiang Gu, Chris Tensmeyer, Tong Yu, Changyou Chen, Jinhui Xu, and Tong Sun. 2022. TiGAN: Text-Based Interactive Image Generation and Manipulation. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 3 (Jun. 2022), 3580–3588. doi:10.1609/aaai.v36i3.20270

A Prompts and Example Captions for Augmentation with Haptic Labels

Prompt	Example Output
`Write a tactile expression description sentence in third person perspective action.`	The ground shudders beneath her feet as the airplane lands.
`Write a tactile expression description sentence including the following attributes.`	The airplane's landing produces a jolt, vibrations, and gentle tremors.
`Write a single sentence that summarizes tactile feedback with the following attributes. Don't write anything about sound characteristics.`	The airplane's landing produces a jolt, vibrations, and tremors.
`Write a tactile expression description sentence including the following attributes. Creative paraphrasing is acceptable.`	The airplane's landing creates a low-frequency hum and gentle thrum.

Table 2: Prompts and example caption variants generated for the input "Large commercial airplane landing at an airport runway". Each prompt is designed to create a tactile-focused description that more closely aligns with what haptic designers may prompt the final model for. Each prompt is provided to the model with a "system" header, and results are expected as JSON delimited string, which reduced "noisy" results with additional non-label text.

B Cross-Entropy Loss With Negative Samples

The cross-entropy loss for positive and negative samples was defined as follows. Let N_{pos} and N_{neg} denote the number of positive and negative samples, respectively, and T the number of timesteps. The number of classes is represented by C. For positive samples, $y^{\text{pos}}_{i,j,k}$ is the true target at sample i, timestep j, and class k, and $\hat{y}^{\text{pos}}_{i,j,k}$ is the corresponding predicted probability. Similarly, for negative samples, $y^{\text{neg}}_{i,j,k}$ and $\hat{y}^{\text{neg}}_{i,j,k}$ represent the true and predicted values, where the predicted logits are flipped as $1 - \hat{y}^{\text{neg}}_{i,j,k}$. The weight λ scales the contribution of the negative samples and is set to 0.1 for all models. The loss is then given by:

$$\mathcal{L}_{\text{CE}} = \left(-\frac{1}{N_{\text{pos}} \cdot T} \sum_{i=1}^{N_{\text{pos}}} \sum_{j=1}^{T} \sum_{k=1}^{C} y^{\text{pos}}_{i,j,k} \cdot \log\left(\hat{y}^{\text{pos}}_{i,j,k}\right) \right)$$
$$+\lambda \left(-\frac{1}{N_{\text{neg}} \cdot T} \sum_{i=1}^{N_{\text{neg}}} \sum_{j=1}^{T} \sum_{k=1}^{C} y^{\text{neg}}_{i,j,k} \cdot \log\left(1 - \hat{y}^{\text{neg}}_{i,j,k}\right) \right)$$

C Questionnaires for survey and Semi-structured Interview

Here is the full list of the questionnaires that we used during the user study. All survey questions follow a 5-point Likert scale from "Strongly disagree" (1) to "Strongly agree" (5).

C.1 Survey

- Factors of Haptic Experience
  - Autotelics 1 : The haptic feedback felt satisfying
  - Autotelics 2 : I like how the haptic feedback itself feels, regardless of its role in the scene
  - Autotelics 3 : I disliked the haptic feedback (*)
  - Autotelics 4 : I would prefer the scene without the haptic feedback
  - Expressivity 1 : The haptic feedback all felt the same (*)
  - Expressivity 2 : I felt adequate variations in the haptic feedback
  - Expressivity 3 : The haptic feedback helped me distinguish what was going on
  - Expressivity 4 : The haptic feedback changes depending on how things change in the system
  - Expressivity 5 : The haptic feedback reflects varying inputs and events
  - Realism 1 : The haptic feedback was realistic
  - Realism 2 : The haptic feedback was believable
  - Realism 3 : The haptic feedback was convincing
  - Realism 4 : The haptic feedback matched my expectations

  (*) reverse-coded the negatively phrased items when analyzing the study results
- System Usability
  - Goal : I was able to achieve what I had in mind with this [model].
  - Iteration : I was able to iteratively improve the output of my generation.
  - Workload : The workload has been improved through [model].
  - Future Use : I would like to use [model] in future tasks as well.

C.2 Semi-structured Interview

- Q: What aspects of the model worked effectively? What could be improved?
- Q: How did you refine your prompts throughout the study?
- Q: Which theme did you find most engaging, and why?
- Q: What element of this framework surprised you?

D Dataset Datasheet for HapticGen

We followed the datasheets for dataset template [19] to document the dataset collected through HapticGen's user studies.

D.1 Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to support the development of HapticGen, a generative model designed to create vibrotactile haptic signals from textual descriptions. This dataset addresses a key gap in the field of haptic design: the lack of large, labeled haptic datasets, which significantly limits the training of machine learning models for this domain. Existing libraries, such as VibViz, contain only a few hundred signals, falling far short of the scale seen in state-of-the-art audio and video datasets, which often comprise hundreds of thousands of examples. By filling this gap, the dataset enables the training of a generative model capable of producing diverse, nuanced haptic signals tailored to textual prompts, thus advancing the accessibility and scalability of haptic design.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

This dataset was created by Youjin Sung, Kevin John, Sang Ho Yoon, and Hasti Seifi. The authors are affiliated with Arizona State University and Korea Advanced Institute of Science and Technology (KAIST).

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the Global Research in the Digital Field program supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) (RS-2024-00419561), the National Research Council of Science & Technology (NST) grant by the Korea government (MSIT) (CRC21014), a research grant from VILLUM FONDEN (VIL50296), and a research grant from the National Science Foundation (NSF) (#2339707).

D.2 Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

Each instance in the data is a tuple of (text prompt, haptic signal, and human vote). Every instance was prompted and rated by the same person. We separate the data into that collected from our intermediate evaluation study, labeled as "expert-voted", which was

included in our model training for HapticGen, and the data from our final evaluation study, labeled as "user-voted" which was not used in training HapticGen.

How many instances are there in total (of each type, if appropriate)?

There are a total of 4526 tuples in the dataset, including the expert-voted dataset with 1297 tuples and the user-voted dataset with 3229 tuples.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified.

The dataset is a curated sample rather than an exhaustive set of all possible text prompts and vibration signals. The expert-voted dataset was created during our intermediate evaluation, where 15 haptics experts provided prompts and rated the generated haptic signals. The user-voted dataset was collected with 32 participants, including both haptics experts and novice users, who generated prompts and rated corresponding signals across various themes.

While the dataset is not representative of every possible text-haptic pairing, care was taken to ensure diversity in text prompts and haptic signals. For instance, prompts were drawn from diverse themes such as sports, interactions, emotions, games, and simulations, which align with common use cases in haptic design. Additionally, expert-voted signals focus on professional preferences, whereas user-voted signals capture a broader range of subjective evaluations, providing complementary perspectives.

Representativeness was not quantitatively validated due to the exploratory nature of the dataset. However, qualitative diversity was emphasized by involving participants with varying levels of expertise and focusing on distinct design scenarios to ensure a wide coverage of potential haptic applications.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

Each instance has a vibrotactile wav file (8 kHz 8 bit PCM) and a JSON metadata file. The metadata file contains the original prompt and user vote.

The metadata structure is as follows:

- **filename**: The name of the haptic signal .wav file generated for the user.
- **user_prompt**: The original natural language description of the tactile sensation provided by the user.
- **model**: The model used to generate the signal.
  The possible models include:
  - "HapticGen": The final, fine-tuned model.
  - "Baseline-AudioGen": A baseline model for A/B testing.
  - "Initial": An early version of the HapticGen model without fine-tuning.
- **vote**: User feedback on the quality of the generated signal, where 1 indicates the generated vibration matches with the prompt (thumbs up), and -1 indicates the generated vibration does not match with the prompt (thumbs down).
- **prompt_variant**: (optional) A variant of the original prompt provided to the generative model for increasing variance of the generated vibration. This prompt variant was not shown to users (not considered for vote).

Is there a label or target associated with each instance? If so, please provide a description.

Yes, each instance in the dataset includes a target in the form of a human vote (thumbs up or thumbs down) indicating the perceived quality or suitability of the generated haptic signal for the corresponding text prompt. These votes are intended to guide model fine-tuning or evaluation.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Participant IDs and background information have been removed for anonymity.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

All instances are independent and can be used as standalone entries. However, instances generated as part of the same batch can be identified, as they share a unique batch ID within their filenames.

Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

We do not provide recommended data splits.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description

Participant prompts and votes are inherently subjective and may reflect individual differences in interpretation or preference for haptic signals. Additionally, some participant prompts may be similar or may result in similar haptic signals, leading to potential overlap within the dataset.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?

The dataset is self-contained.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor– patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.

No, the dataset does not contain any data that might be considered confidential. All participant data has been anonymized.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

The dataset consists of participant-generated prompts and associated haptic signals, which were created during the course of user studies. Some scenarios (e.g., shooting a gun) may describe intense or evocative situations, such as those in video games. These prompts reflect the creative input of the participants and are included to capture a wide range of potential haptic use cases.

Does the dataset identify any subpopulations (e.g., by age, gender)?

No, all data was anonymized to focus solely on the haptic signals, text prompts, and associated votes.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?

No, it is not possible to identify individuals either directly or indirectly from the dataset. All identifying information, such as participant IDs, has been removed to ensure anonymity, and the dataset does not include any personal or demographic information that could link data points to specific individuals.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?

The dataset does not contain sensitive personal data linked to participants, such as race, ethnicity, or other identifying information. While some participant-generated prompts may include fictional scenarios that reference demographic details (e.g., age, nationality, or location), these are not tied to any individual and are purely narrative style for the purpose of haptic design.

D.3 Collection Process

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

The data associated with each instance was acquired during multiple user studies. The text prompts were reported directly by participants during user studies, reflecting their creativity and intent for generating haptic signals. The associated haptic signals were generated by a model in response to these prompts, and participants provided direct ratings (thumbs up or thumbs down) to evaluate the suitability of the generated signals. While the text prompts and ratings were not externally validated due to their subjective nature, the data collection process (i.e., our evaluation studies) ensured that the dataset captures a broad range of haptic themes and user preferences and perspectives from both expert and novice haptic designers.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?

The data was collected during user studies using a combination of manual and automated mechanisms. Participants entered text prompts through a custom desktop graphical user interface developed for the HapticGen system. The HapticGen model generated vibrotactile haptic signals in response to these prompts, which were played back on Meta Quest Virtual Reality (VR) controllers via an OpenXR application integrated with the desktop interface. Participants felt the vibration on the VR controllers and evaluated the suitability of the generated signals by providing thumbs up or thumbs down ratings through the same interface. The mechanisms and procedures were validated through iterative pilot testing to ensure usability, functionality, and reliability in the data collection workflow. The use of standard commercial hardware for haptic playback, such as Meta Quest VR controllers, ensured consistent and reliable rendering of the generated haptic signals in a practical and accessible manner.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

The dataset can be viewed as a curated and rated subset of all possible text prompts and vibration signals. The text prompts were decided by participants during the studies, reflecting their creativity and intent, while the vibration signals were generated by an autoregressive transformer-based model in response to those prompts.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

The data collection process involved participants recruited from academic and professional communities, including professors, researchers, students, and industry professionals from VR development teams. Compensation for each study was based on the estimated study time. It is set at USD 7.5 per hour exceeding the local minimum wage. All compensation was provided in cash and distributed following approval from the Institutional Review Board (IRB).

Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

All data was created and collected in the spring and summer of 2024.

Were any ethical review processes conducted (e.g., by an institutional review board)?

Yes, the study was approved by the institutional review board (IRB) at the authors' university.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

We collected the data directly through in-person user studies.

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

Yes, this research project was performed under approval from (Korea Advanced Institute of Science & Technology (KAIST)'s) IRB (KH2023-187). We followed the form provided by the institution (e.g., Informed Consent for Human Subjects). The notification was provided through the KAIST Institutional Review Board Form 4, which details the purpose, scope, and retention of personal data, as well as participants' rights. Below is the exact language of the key notification; Consent to the collection and use of personal information(Purpose, Collected Data, Retention Period, and Disadvantages of Refusal), Sharing of personal information with third parties, and Consent options).

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

Yes, participation was voluntary. Participants reviewed and signed the consent form at the beginning of the data collection. KAIST Institutional Review Board Form 23 is used to obtain participant consent for collecting personal information necessary to process compensation for participating in a research project.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

Yes, participants could revoke their consent during the study and prior to the publication of anonymized data. The name and phone number of the researcher were provided in the consent form in case participants needed additional information or if they wanted to revoke their consent.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

No formal data impact analysis was conducted. However, the dataset is fully anonymized and any identifiable data is removed to minimize impact on and privacy risks to participants.

D.4 Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description.

Instances with incomplete data, such as missing ratings, were removed.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

We do not provide the raw data at this time but may in future. If so, this data will made available through the same OSF and GitHub repositories.

Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.

No, we do not provide this code.

D.5 Uses

Has the dataset been used for any tasks already? If so, please provide a description.

The expert-voted dataset was used to fine-tune the HapticGen model, as described in Section 7. The user-voted dataset, while suitable for similar use, has not yet been utilized for this purpose.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

No. We have not created such a repository.

What (other) tasks could the dataset be used for?

As the largest publicly available labeled vibration dataset to date, this resource can be useful for various applications in haptics research and design. Beyond supporting text-to-vibration models, the dataset could facilitate tasks such as haptic captioning, where the objective is to generate textual descriptions from haptic signals. Additionally, it could serve as a comprehensive vibrotactile effect library, enabling rapid search, recommendation, and integration of haptic feedback for various end user applications.

Are there tasks for which the dataset should not be used? If so, please provide a description.

While the dataset is suitable for many haptic research tasks, we recommend that users consider its limitations, such as the fact that it is not exhaustive in its coverage of possible vibrotactile signals or prompts. Nonetheless, it remains a valuable starting point for exploring diverse applications and should be supplemented with additional data sources as needed.

D.6 Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

Yes, the dataset will be made open-source and publicly available to anyone for research and development purposes. It is intended to support the broader scientific community in advancing haptic design and related fields.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?

The dataset will be available on GitHub at https://github.com/HapticGen/hapticgen-dataset and through the Open Science Framework (OSF) at https://osf.io/vdmej/ which is assigned a digital object identifier: DOI 10.17605/OSF.IO/VDMEJ.

When will the dataset be distributed?

At the same time as the publication of the paper.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to.

The dataset is released under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0): https://creativecommons.org/licenses/by-nc/4.0/.

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No.

D.7 Maintenance

Who will be supporting/hosting/maintaining the dataset?

The authors of this paper will host and maintain the data for at least three years after public release.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

The authors can be contacted at 672@kaist.ac.kr, kevin.john@asu.edu, sangho@kaist.ac.kr, hasti.seifi@asu.edu.

Is there an erratum?

This information will be posted on GitHub.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?

Yes, the authors will update the dataset if needed and publish the changes via GitHub for at least three years after the public release of the dataset.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.

No specific limits on data retention were established, as the dataset was anonymized and did not include personally identifiable information.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.

Yes, historic versions of the dataset will remain available through GitHub and the Open Science Framework (OSF).

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.

Yes, we welcome others to fork the dataset via GitHub. We will actively validate and accept pull requests to link to such contributions from our dataset repository.