

May the Force Be with You: Dexterous Finger Force-Aware VR Interface

Fengze Zhang*[†]

Yunxiang Zhang*[‡]

Xi Peng

Sky Achitoff

Paul M. Torrens

Qi Sun

New York University

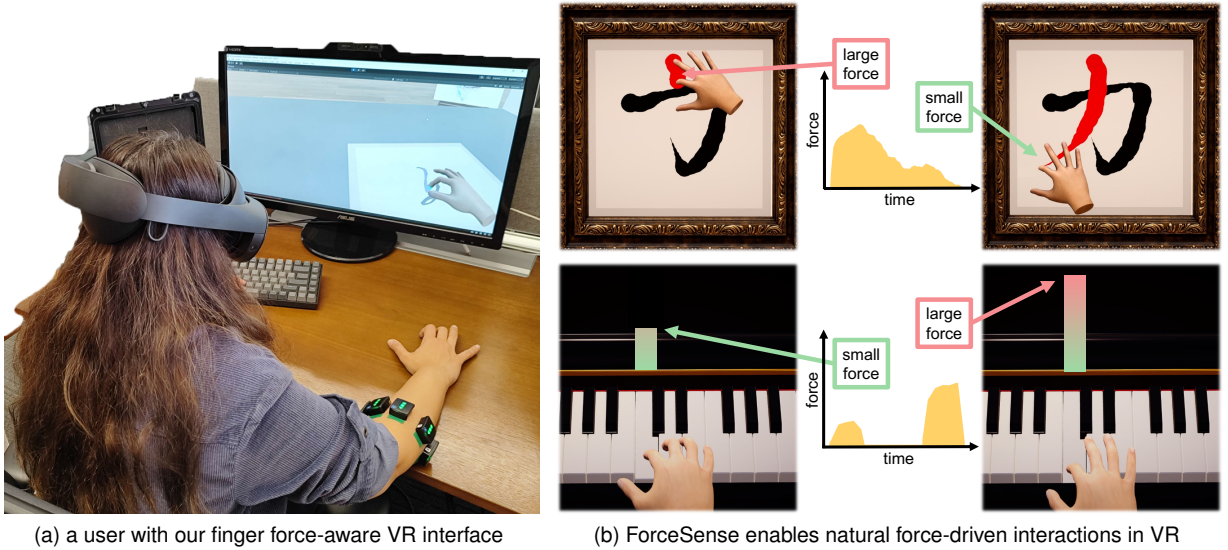


Figure 1: *Our finger force-aware VR interface, ForceSense*, (a) offers a natural way to perform force-driven interactions, and (b) supports a variety of VR applications that require high-finesse skills, such as calligraphy (top) and piano playing (bottom).

ABSTRACT

Advances in virtual reality (VR) have reduced experience differentials for users. However, gaps between reality and virtuality persist in tasks that require coupling users’ multimodal physical skills with virtual environments in delicate ways. User embodiment in VR easily breaks when physicality feels inauthentic, especially when users invoke their innate predilection to touch and manipulate things that they encounter. In this research, we examine the potential of force-aware VR interfaces for enabling natural connections to user physicality and evaluate them in high-finesse cases of touch. Combining surface electromyography (sEMG) with visual tracking, we develop an end-to-end learning-based system, ForceSense, to decode users’ dexterous finger forces from their forearm sEMG signals for direct usage in standard VR pipelines. This approach eliminates the need for hand-held tactile equipment, thereby promoting natural embodiment. A series of user studies on manipulation tasks in VR validate that ForceSense is more accurate, robust, and intuitive than alternative solutions. Two proofs-of-concept VR applications, calligraphy and piano playing, demonstrate that the good synergy between visual, auditory, and tactile modalities, as ForceSense affords, has the potential of enhancing users’ task learning performance in VR. Our source code and trained models are released at <https://github.com/NYU-ICL/vr-force-aware-multimodal-interface>.

Index Terms: Force sensing, tactile perception, sEMG learning.

*Both authors contributed equally to this research.

[†]E-mail: fz2244@nyu.edu

[‡]E-mail: yunxiang.zhang@nyu.edu

1 INTRODUCTION

VR systems are increasingly capable of matching the physicality of user behaviors to counterpart simulated media. Recently, this has culminated in the development of what is termed reality-to-simulation (Real2Sim) [23]. The key promise of Real2Sim is to match virtual phenomena to their real-world counterparts with significant parity in behaviors, underscoring broader visions for digital twins and other simulation technologies. The potential for Real2Sim to bring humans into this ecosystem is significant, and VR is currently the most promising media for doing so. However, VR-based Real2Sim efforts are artificially constrained, in particular, for high-finesse tasks, i.e., those requiring delicate physical abilities of dexterity and tactile manipulation. Available options such as handheld controllers and vision-based hand tracking are too blunt to reach significant parity with human experiences. Here, we explore the notion of fine-resolution embodiment between users’ natural behaviors to reach out for, grasp, manipulate, and explore simulated objects in VR using their natural proclivities of dexterity and matching sense of tactility in the environment.

We demonstrate the case of producing force-awareness in VR systems as a particular opportunity for advancing Real2Sim interfaces. Users’ sense of force taps into a crucial pathway for visual-auditory-tactile connection among users, objects in virtual form, and simulated phenomena of object action and interaction with the user ego-self [24, 4, 13, 57]. This triple, we consider, is accessible through the use of muscle-computer interfaces and we show one such pathway via commercially-available surface electromyography (sEMG), which we harness as a *direct* interface between user dexterous behavior and virtual content.

As we show, a snug multisensory “fit” must be established between human dexterity while interacting with the physical environment and simulated response. To equip VR systems with tac-

tile sensing and interfaces to narrow the Sim2Real gap, researchers have explored various ways to acquire tactile forces and generate haptics, ranging from wearables [49, 34] and touch surfaces [13] to ultrasonic haptic rendering [31], vibrotactile gloves [64, 68], and electro-tactile stimulation [50]. However, due to their requirements of being handheld, natural user tendencies to use their fingers and thumbs are often interrupted, which can result in a flatter of their embodiment while on-task. Nonetheless, sEMG has emerged as a promising technique for designing low-friction VR interfaces [30, 51]. In particular, prior work has successfully demonstrated the robust decoding of users’ finger-applied forces with forearm-mounted sEMG sensors used alone [66, 61].

To this end, we propose a novel paradigm of force-aware human-environment interactions for VR. This is achieved by inferring VR users’ fingertip forces from sEMG data and applying them to drive application-specific physical simulations in real time. Importantly, our approach enables users to interact with the virtual environment in a natural bare-hand manner without compromising their dexterous finger movements in tasks requiring high-finesse skills.

In particular, we introduce ForceSense, a robust yet lightweight transformer-based sEMG-to-force prediction model. It decodes finger forces in Newton from sEMG-captured forearm muscular activities during hand-object interactions. Together with camera-tracked hand/finger poses, the decoded forces are mapped as inputs to drive physically based simulations in virtual environments.

We conducted objective evaluations to validate the accuracy and robustness of ForceSense, as well as user studies to demonstrate its effectiveness in force-driven elemental interactions under a natural bare-hand setup. Moreover, through two case studies of multimodal task learning, piano playing and calligraphy, we show that the good synergy between the visual, auditory, and tactile senses of VR users, as ForceSense enables, has the potential of boosting their sense of presence and task performance in VR. This is evidenced by a 32.1% (piano) and an 11.1% (calligraphy) task performance improvements and a 76.2% overall user preference compared to a hand-tracking-based interface devoid of the tactile modality.

We hope this research will provide insights and motivate a novel frontier for developing multimodal VR interfaces toward physically realistic and natural embodied user experiences.

In summary, we make the following main contributions:

- ForceSense, an accurate, robust, and lightweight finger-level sEMG-to-force prediction model;
- A ForceSense-enabled VR interface that supports dexterous force-driven hand-object interactions in VR;
- A series of objective and subjective evaluations of ForceSense, as well as two applications, virtual calligraphy and piano playing.

2 RELATED WORK

A number of existing VR interfaces provide the foundation for our work. Immersive interfaces have been explored in VR [68], for example, showing variation in input modality between held controllers and free-hand interaction for physical exertion, agency, task performance, and motor behavior, e.g., for mid-air touching, ray-casting, selection, and trajectory-tracing tasks. While using VR controllers, users reported less physical exertion, felt more in control, and were faster and more accurate [7, 21, 19, 35].

Tracking technologies present another interface approach. For example, high-resolution (video-based) hand-tracking has been shown to support natural gesture-based control [62] and remote pointing by ray-casting [27]. Eye-tracking has also been used to drive hands-free interactions in VR [43]. Moreover, multimodal interfaces have been demonstrated [18], as supporting embodiment in one-way interactions to virtual environments [52] as well as for

two-way relationships in an expanded form [9]. Amid these exciting developments, tactile sensing, as a core modality of human-environment interaction, remains under-investigated. This is perhaps a significant remaining omission for VR and Real2Sim, which remains challenging due to the intricacy needed to sense contact forces without influencing natural task dexterity.

A large body of prior literature has focused on generating haptic sensations and introducing them to VR users, commonly referred to as *haptic rendering*. This includes actuator-based [64], ultrasound-based [31], and electrical muscle stimulation [50, 32, 33] devices to create haptics from physical simulation. Using wearable electronics that detect point-wise contacts [37, 45], users can rapidly type [48], grasp [13], or avoid collision [53] in VR. Similarly, wearable force-sensing interfaces, including gloves and knitted surfaces, have been developed to realize force-driven interactions and haptic perception [47, 49, 34]. While these interfaces enable force-awareness in VR, their bulky on-hand design largely constrains users’ dexterous finger movements in tasks requiring high-finesse manipulation skills and thus inevitably compromises natural user embodiment.

Haptic and touch approaches of course have advantages over controllers in their ability to garner physicality sensations, but they are often *orthogonal to the scope of this research* if the intent is to evoke a two-way dynamic between the user and VR content. Indeed, a reality gap can easily open-up if users “feel” objects (as a one-sided interface) without first probing, prodding, pushing, or pinching them. User initiation of the interaction, we contend, is a very significant precursor to embodied interaction, sourced in their marshaling of thoughts and actions in the brain. In other words, how users perceive an object or an interaction is significantly played out in the mind before action is initiated. Convincing response is then, and only then, confirmatory and it is this interplay that evokes authentic embodiment (see [9]).

The widespread commercial availability of sEMG sensors has established some novelty in embracing electromyography as a potential neural interface and an obvious question is whether sEMG could be used to drive VR. sEMG can be unobtrusive, with small sensors placed on-arm that allow weak electrical signals propagating within the muscles to be read and translated into finger-wise activities or general pattern recognition [42]. The approach has, for example, been shown for classifying gestures [10, 17, 59, 20, 51]. If considered in possible concert with deep learning, sEMG can continuously track hand poses [30]. This opens up very significant Real2Sim pathways, including vicarious robotic control [14]. More broadly, compounds of sEMG signals could infer human-environment contact forces, with potential use for hand gripping [16, 61], and finger touching [11, 6, 66]. Prior work has leveraged deep neural networks to model the mapping between sEMG signals and such force information [36, 5, 66]. Despite the possibility of decoding hand/finger-level forces from forearm sEMG signals, we still have little knowledge of how immersive VR applications would benefit from such an additional tactile modality and how it adds to VR users’ multimodal experience in the virtual environments. To our knowledge, empirical evaluations of task performance relative to traditional controller-based interfaces have not been done.

3 METHOD

3.1 Data Collection

Anatomical studies of human upper limbs have revealed that the major muscle groups controlling finger-level force/torque exertions and the majority of hand movements are situated on the forearms [12, 29]. To capture the activities of the relevant muscles and map them to finger forces, we attached 8 Delsys Trigno wireless surface electromyography (sEMG) sensors [1] to the forearms of the user. The mounting location was chosen to be close to the upper end of the forearm (approximately 4-6cm to the elbow), where the sEMG signal strength is the strongest. As shown in Figure 2, the 8 sEMG

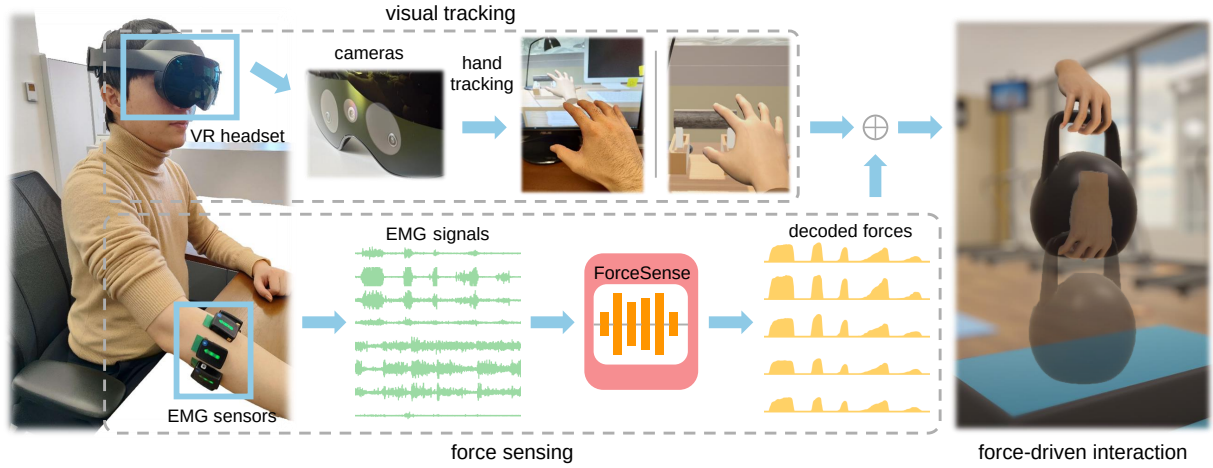


Figure 2: *Workflow of our dexterous finger force-aware VR interface.* ForceSense provides fast, robust, and accurate finger force sensing that, together with visual tracking, enables VR users to accomplish complex interaction tasks requiring fine-level dexterity in virtual environments.

sensors formed a uniformly spaced circular layout. To measure the user’s finger-exerted forces, we used a Sensel Morph force-sensing trackpad [2] and let the user apply forces to it with the appropriate finger combinations for the interaction task. To obtain synchronized forearm sEMG and finger force data, we prepared a multi-threading Python program to synchronize wirelessly streamed sEMG signals (sampled at 2000Hz) with sensed force data (sampled at 120Hz).

3.2 Data Pre-processing

As is typical for bio-electrical signals, prior literature has suggested that raw sEMG recordings contain a non-trivial amount of electromagnetic artifacts and measurement noises [44, 41, 67] and is not an appropriate representation for machine learning models to work with. Following recent research that deployed deep neural network models to sEMG data [65, 8, 66], we converted raw sEMG signals from the time domain to the frequency domain by applying short-time Fourier transform (STFT). For each sEMG channel, we used a Hanning window of length 256 and a hop length of 32 to transform a 1-dim sEMG sequence into a 2-dim sEMG spectrogram with 129 frequency bins. The resulting sEMG data representation has a size of (C, F, T) , where $C = 8$, $F = 129$, and T represent the number of sEMG channels, frequency components, and frames in the spectrogram. Note that the original sEMG signals were also downsampled from 2000Hz to 62.5Hz during this process. To temporally align the force data to the processed sEMG spectrograms, we performed the nearest neighbor interpolation to downsample the raw finger force data samples to 62.5Hz.

3.3 ForceSense: Finger-Level sEMG2Force Prediction

To establish a robust mapping from forearm sEMG signals to finger forces, we resort to the supervised learning paradigm and leverage a deep neural network to learn the mapping from data. Particularly, we adopted two prominent architectures: convolutional neural networks (CNNs) [26] and transformers [54]. CNNs excel at identifying patterns and extracting features from structured data, such as the 2D spectrograms we have. Transformers, on the other hand, are adept at processing sequential data. The self-attention mechanisms are efficient in extracting contextual information and relationships within sequences. When merged together, these architectures offer complementary advantages for interpreting the sEMG spectrogram sequences toward finger-level force decoding.

Model architecture and data flow. The inputs to our model ForceSense is a 4D tensor of shape (B, C, F, T) , where B , $C = 8$, $F = 128$, and $T = 32$ represent the batch size, number of sEMG

channels, number of retained frequency components, and spectrogram sequence length. Notice that we drop the highest frequency component to make the number of frequency bins a power of 2. At the bottom of ForceSense, we adopt a CNN model to extract deep features from the multi-channel sEMG spectrogram sequences. The CNN model consists of 4 repeating Convolution-BatchNorm-ReLU blocks, with the first 3 blocks each followed by a 4×1 Maxpool layer and the last block followed by a 2×1 Maxpool layer. As a result, the frequency dimension is downsampled from 128 to 1 and subsequently squeezed out. All the convolutional layers leverage a 3×3 kernel size.

Given the CNN outputs, a 3D tensor of shape (B, C, T) , we use a transformer layer to capture the temporal dynamics and relationships across CNN-extracted deep features and estimate the finger forces corresponding to the input sEMG signals. This transformer model comprises 10 transformer encoder blocks, each containing eight attention heads. The outputs is a 3D tensor of shape (B, J, T) , where B , $J = 5$, and $T = 32$ represent the batch size, number of force components (i.e., number of fingers), and the length of output force sequence. Notably, we make ForceSense take in sEMG spectrogram sequences which are 32 frames long (equivalent to 624ms) for taking advantage of relevant muscle activity information from the near past. Only the predicted force values contained in the most recent frame of the 32-frame outputs are kept and used throughout our experiments. At the inference stage, we use a sliding window to slice the most recent part of the sEMG data stream and feed it to ForceSense to make predictions on the fly.

Loss function. Given the input sEMG data stream, ForceSense should be able to determine the specific finger or fingers from which the forces originate as well as the magnitude of the forces. Following a recent work on sEMG-based force decoding [66], we leverage a sigmoid function to convert the raw outputs of the transformer model in ForceSense into the likelihood of fingers applying forces at each time frame. This sequence of likelihood values is then compared against the ground truth through a cross-entropy loss:

$$L_c = \frac{1}{T \cdot J} \sum_{t=1}^T \sum_{j=1}^J g_j^t \cdot \log p_j^t + (1 - g_j^t) \cdot \log(1 - p_j^t) \quad (1)$$

where $p_j^t \in [0, 1]$ and $g_j^t \in \{0, 1\}$ represent the predicted likelihood and ground truth of the j_{th} finger exerting force at the t_{th} time frame.

The finger force is computed as $\hat{F}_j^t = 2F_{max} \cdot \max(0, p_j^t - 0.5)$, where F_{max} denotes the pre-defined force range. An L_2 loss is uti-

lized to supervise the force regression:

$$L_r = \frac{1}{T \cdot J} \sum_{t=1}^T \sum_{j=1}^J \|\hat{F}_j^t - F_j^t\|^2 \quad (2)$$

Finally, we trade off the classification and regression objectives through a weighting term λ . Throughout our experiments, λ is set to 4 if not explicitly specified otherwise.

$$L_{\text{total}} = L_c + \lambda \cdot L_r \quad (3)$$

4 EVALUATION

To comprehensively validate our force-aware interface ForceSense, we first quantitatively evaluate its predictive power and system performance by establishing comparisons with prior arts (Section 4.1), then conduct a user study where ForceSense is applied to facilitate several force-driven virtual object interactions (Section 4.2). Using ForceSense, we create two practical VR applications, virtual piano playing and calligraphy practicing, and analyze users’ task learning and completion performance (Section 4.3).

4.1 Real-Time Finger Force Decoding with ForceSense

Prior research has explored using machine learning tools to model the mapping from forearm sEMG signals to hand/finger-level forces in a data-driven manner [6, 36, 11, 16, 66]. To better position our transformer-based predictive model ForceSense in the context of prior literature, we perform quantitative evaluations of its force decoding performance on a public EMG2Force benchmark dataset [66] and establish comparisons against a recent autoencoder-based EMG2Force model [66]. The dataset contains 148.5-minute time-synchronized forearm sEMG and finger-wise force data collected from 9 subjects. It uniformly covers 11 frequently used finger combinations during hand-object interaction, with 6 pressing/pushing-like actions and 5 pinching/gripping-like actions. From the thumb to the pinky finger, the force level in Newton goes up to 29.8, 24.4, 25.6, 20.4, and 14.7, respectively.

Evaluation metrics. Following prior literature on finger-level force estimation [36, 5, 66], we adopted the following quantitative metrics for predictive performance evaluation: classification accuracy, normalized root mean squared error (NRMSE), and coefficient of determination R^2 . Note that for the classification metric, we follow previous work [66] and only count the model’s predictions as correct if all five fingers are correctly classified.

Experimental setup. The train-test split described in [66] was followed for training and evaluation. We implemented ForceSense using the PyTorch deep learning framework [39] and optimized it using an Adam optimizer [22] against the loss function defined in Equation (3) for 100 epochs. A batch size of 256 was employed. The learning rate started at $1e^{-4}$ and was multiplied by 0.7 every 10 epochs. All hyper-parameters described above were determined via a cross-validation procedure.

Model performance. As shown in Table 1, ForceSense largely outperforms the baseline model across all three evaluation metrics, showing an average accuracy, NRMSE, and R^2 of 91.35%, 3.28%, and 89.59%. Notably, such predictive performance is obtained with a modest model size of 25.2M trainable parameters only. Given an 8-channel sEMG data stream sampled at 2000 Hz, ForceSense only requires 793.98M MACs (multiply-accumulate operations) to make an inference step of batch size 1. This computation on average takes 5.50ms on a desktop PC equipped with an Nvidia RTX 4090 GPU, i.e., ForceSense can roughly run at 180 FPS on the same device.

Table 1: Comparison of EMG2Force predictive performance. The $\uparrow/\downarrow$ signs denote that higher/lower scores mean better performance.

Method	Accuracy $\uparrow$	NRMSE $\downarrow$	R^2 $\uparrow$
[66]	88.83 $\pm$ 6.13%	3.29 $\pm$ 1.76%	85.82 $\pm$ 14.96%
Ours	91.35$\pm$1.83%	3.28$\pm$0.58%	89.59$\pm$2.60%

Discussion. Prior literature has suggested that low system latency and high refresh rate are indispensable components to building successful VR applications. In particular, users’ motor performance, sense of agency, and simultaneity perception may be negatively affected when the system response delays exceed 50 ms [56, 3]. Empirical studies on the impact of system refresh rate on VR users’ task performance and simulator sickness level revealed that 120 FPS is an important threshold beyond which no negative effects from low system refresh rate can be observed [25, 58]. Notably, ForceSense runs at 180 FPS and has a system delay of 7ms, satisfying the requirements described above by a large margin. In addition, ForceSense’s average force decoding performance in terms of NRMSE evaluates to 3.28%, equivalent to an absolute error of 0.98N only. These results evidence ForceSense’s practicality for real-time VR applications.

4.2 Force-Driven Interaction with Virtual Objects

Leveraging ForceSense’s force decoding capability, we designed a user study that consists of 3 hand-object interaction scenarios commonly encountered in VR (Figures 3a to 3c). The goal is to study to what extent ForceSense improves users’ perception of the *physical workload* required by different virtual interaction tasks.

Participants. We recruited 8 subjects (ages 22-26, 3 female) to join the study. All have normal or corrected-to-normal vision, no history of auditory deficiency, and no finger/hand/arm impairments. Among the participants, 4 had no prior experience with VR, 2 used VR equipment 5 times or less, and 2 are familiar with and have regular access to VR devices. None were aware of the hypothesis, the research, or the number of task difficulty levels. The study protocol was approved by the Institutional Review Board (IRB) at the host institution and all subjects gave informed consent before the study.

Calibration. It is a well-known fact that muscle-to-sEMG patterns may exhibit large variations across individuals with different physiology. As a result, machine learning models trained on existing forearm sEMG datasets have significant difficulty generalizing to new users and arm/hand postures [42, 66]. To ensure that ForceSense maintains consistent force decoding accuracy throughout the study, we conducted a pre-study calibration to build a user-specific ForceSense model for each subject. Similar to the procedures described in Section 3.1, we recorded 2-minute forearm sEMG and finger force data with each subject while they performed each elemental interaction shown in Figures 3a to 3c. A hard plastic cuboid box (35cm $\times$ 25cm $\times$ 15cm) fixed on a desk was used as the proxy for subjects to exert forces. Subjects were seated in front of the proxy box, with their elbows (under natural rest state) being at the same height level as the upper facet of the box. They were instructed to pull the back facet, push the front facet, and press the upper facet of the box by naturally extending their arms to realize the three elemental interactions. The Sensel Morph force trackpad was attached to the corresponding facet of the box to measure subjects’ finger-level forces. The total of 6-minute EMG-force data was then used to fine-tune the pre-trained ForceSense model described in Section 4.1 into a customized one. As shown in Figures 3g and 3h, pre-training user-specific ForceSense models on the public dataset [66] largely reduces the necessary data volume for calibration. In addition, our decision to collect 2-minute data

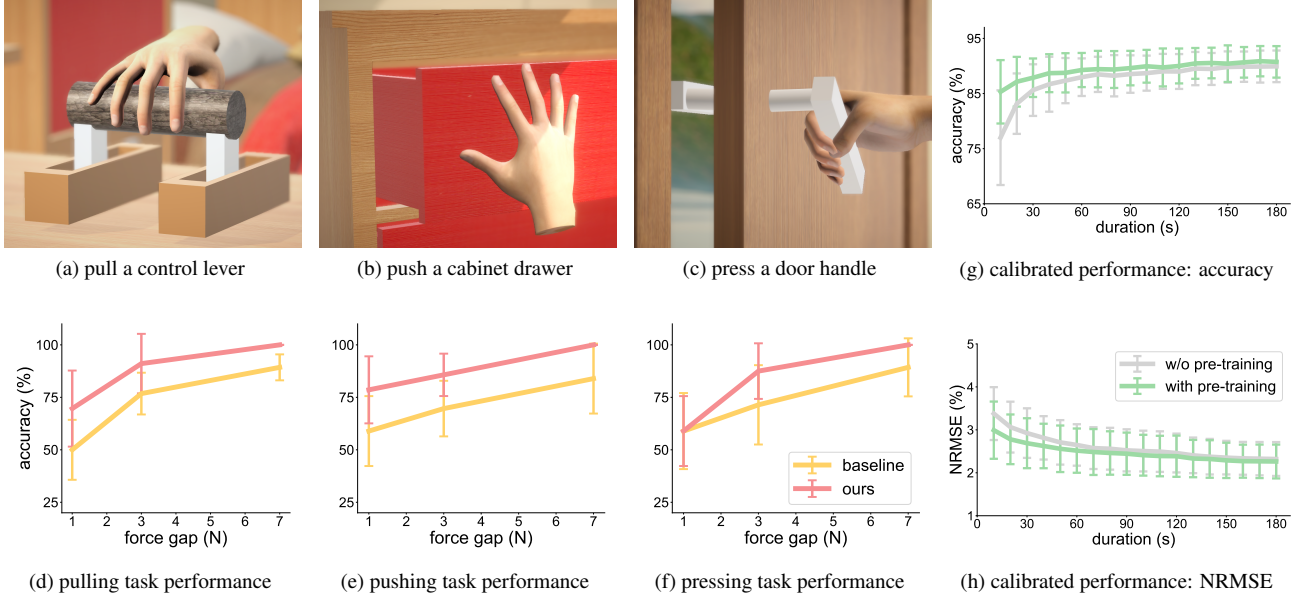


Figure 3: *Elemental interactions with virtual objects via ForceSense.* As presented in (d)/(e)/(f), ForceSense improves users’ performance in accurately perceiving the physical workload required by different virtual interaction tasks as compared to the baseline method.

only for each interaction type was informed by the fact that the model performance gradually plateaus beyond this duration, as can be observed in Figures 3g and 3h.

Conditions. Besides our force-aware interface ForceSense, we included a tracking-based interface commonly used in existing VR platforms as the baseline condition for performance comparison. For ForceSense, we adopted the same proxy box as during the calibration process for subjects to interact with and exert forces. The decoded finger forces were applied to the virtual objects to realize the force-driven interactions in Figures 3a to 3c, i.e., pulling a control lever, pushing a cabinet drawer, and pressing a door handle. For the baseline, we adopted a standard velocity-based approach from prior literature [69] and mapped users’ hand-moving speed (we used the speed of the palm joint tracked by Oculus VR SDK) to varying force levels for the same set of interactions. The speed range [0, 1] in m/s was linearly mapped to the force range [0, 30] in N. Subjects were instructed to apply a desired force intensity level by properly adjusting their hand movement speed. Note that the baseline method cannot exploit the proxy box for improved interaction experience because its interaction mechanism requires an obstacle-free space to generate movement-induced forces.

Tasks. We designed a two-interval forced choice (2IFC) psychophysical task [63] to measure users’ ability to differentiate between virtual interactions requiring different levels of physical workload. Taking the “pulling a control lever” interaction as an example: during each trial of the study, users were instructed to interact with two control levers having different levels of resistance forces, one after another, and indicate which one required more physical effort based on their observations. The resistance force gap between the two control levers was used to adjust the task difficulty levels. The “pushing a cabinet drawer” and “turning a door handle” interaction tasks followed a similar procedure. Both ForceSense and the baseline were employed to complete the same set of 2IFC tasks, and the percentage of trials where users provided the correct answer was reported for performance evaluation.

Setup and procedure. During the study, subjects wore a Meta Quest Pro headset for visuals and audio. 8 Delsys Trigno sensors

were mounted to their right forearm to capture and stream sEMG signals in real time. The VR applications for running the 3 interaction scenarios (Figures 3a to 3c) were implemented using the Unity game engine [38]. The study was split into 3 sessions (one for each interaction), with forced breaks in between to avoid muscle fatigue. A brief warm-up of 12 trials was provided before the start of each session to familiarize participants with the task. Each session had $(\#condition = 2) \times (\#difficulty = 3) \times (\#repetition = 7) = 42$ trials. Each interaction task in a 2IFC trial had a fixed time limit of 10s, with the option of terminating earlier to proceed. For the difficulty levels, we used three force level pairs centered at 25N with a gap of 1N, 3N, and 7N (i.e., 24.5N vs 25.5N, 23.5N vs 26.5N, 21.5N vs 28.5N). To avoid order-induced performance bias, the condition order was counterbalanced across subjects. The order of appearance in each 2IFC pair and for the task difficulty levels was randomized. The study, including hardware setup, pre-study instructions, warm-up sessions, and breaks, took around 2.5 hours per subject.

Results and discussion. All 8 participants’ aggregated 2IFC results are summarized in Figures 3d to 3f. ForceSense consistently outperforms the baseline method in all 3 interaction scenarios. The mean discrimination accuracy with ForceSense is higher than 85% when the force difference is larger than 3N and reaches perfect discrimination for force differences beyond 7N. Note that an absolute force difference of 3N (23.5N vs 26.5N) is equivalent to a relative difference of 12.7% in our study. These results demonstrate ForceSense’s effectiveness in enhancing VR users’ perception of physical workload in VR and highlight the significance of tactile control to immersive VR applications.

4.3 Case Studies: Multimodal Task Learning

In addition to simple force-driven interactions with virtual objects, we further implement two proofs-of-concept VR applications with ForceSense, piano playing and calligraphy (Figures 4a and 4b), to validate its robustness in tasks requiring high-finesse manipulation skills. Through a series of imitation tasks, we evaluate users’ task completion quality and collect their subjective feedback, including perceived usability, physical workload, and sense of presence.

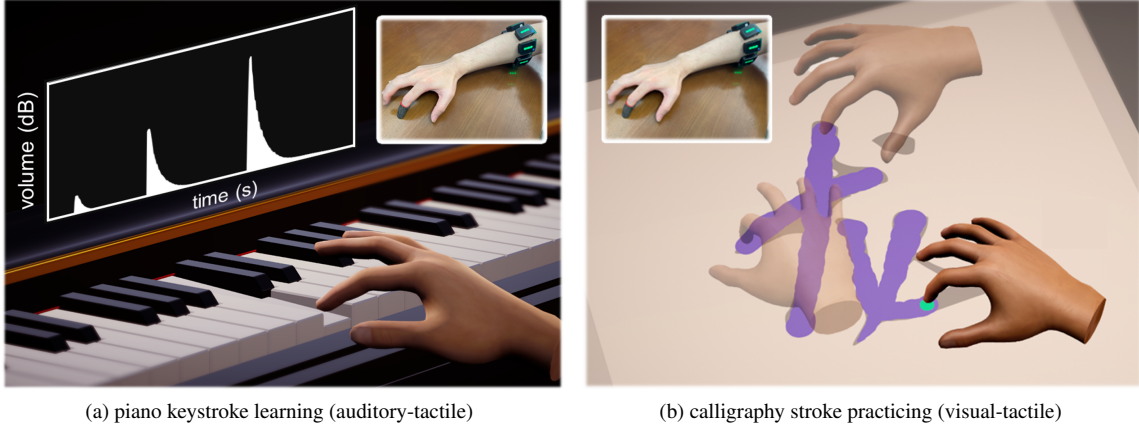


Figure 4: VR live views of tactile-dexterous multimodal task learning with *ForceSense* for (a) piano playing, and (b) calligraphy practicing. Note that the volume plot in (a) is shown for illustration purposes only and was not available to the subjects during the study.

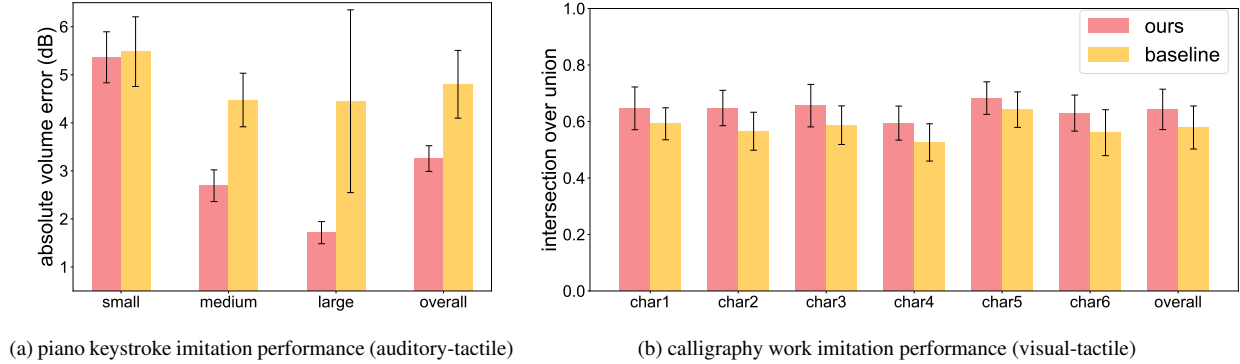


Figure 5: *Objective evaluation of multimodal task learning.* (a) shows users’ auditory-tactile task performance during piano playing. *Lower* volume errors indicate better-imitated piano keystrokes. (b) shows users’ visual-tactile task performance while reproducing calligraphy works. *Higher* IoU scores indicate smaller discrepancies between the reference and reproduced calligraphy works, thus better imitation performance.

Participants. We recruited 12 subjects (ages 20-28, 5 female) to join the study. All have normal or corrected-to-normal vision, no history of auditory deficiency, and no finger/hand/arm impairments. Among the participants, 3 had no prior experience with VR, 4 used VR devices 5 times or less, and 5 are familiar with and have regular access to VR devices. None were aware of the research hypothesis. The research protocol was approved by the Institutional Review Board (IRB) at the host institution and all subjects gave informed consent before the study.

Calibration. Similar to the user study described in Section 4.2, we collected 4-minute EMG-force data from each participant as a pre-study calibration step to customize the pre-trained *ForceSense* model from Section 4.1 into user-specific ones. During the calibration data collection, participants were asked to tap or press on the Sensel Morph force trackpad (fixed on a table) with their index finger only, using varying force levels, hand and finger posture, and contact duration for better data coverage. The resulting customized models were then used to decode users’ finger-level forces in real time for both case studies below.

Experimental setup. During the study, subjects wore a Meta Quest Pro headset for displaying visuals and audio. 8 Delsys Trigno sensors were mounted to their right forearm to capture and stream sEMG signals in real time in the same way as before (Section 4.2).

The scenes and programs for virtual piano playing and calligraphy practicing (Figures 4a and 4b) were implemented in Unity [38].

4.3.1 Auditory-Tactile Task: Piano Keystroke Imitation

Conditions. For performance comparison, we adapted a hand-tracking interaction method similar to the one in Section 4.2 as the baseline. For *ForceSense*, we employed a large table as the physical proxy and mapped a specific area on it to the piano keys. Subjects were instructed to tap on the table as if they were performing piano keystrokes. Their index finger location detected by the cameras on the headset was used to determine the piano key that they intended to strike, and the model-predicted forces were applied to create the keystrokes. For the baseline method, we mapped the moving speed of the user’s index fingertip to varying force levels and used the converted forces to drive the intended piano key. The speed range [0, 70] in °/s was linearly mapped to the force range [0, 17] in N.

Tasks. We designed an auditory imitation task to study users’ performance in reproducing simple piano recordings (we only consider single keystrokes here) by applying an appropriate amount of forces to generate keystrokes. At the start of each trial, a reference piano recording was played. After that, subjects were prompted to imitate the reference recording to get a matching volume profile. Subjects were allowed to perform 5 imitation attempts in each trial. The absolute volume error in decibels was adopted as the metric to

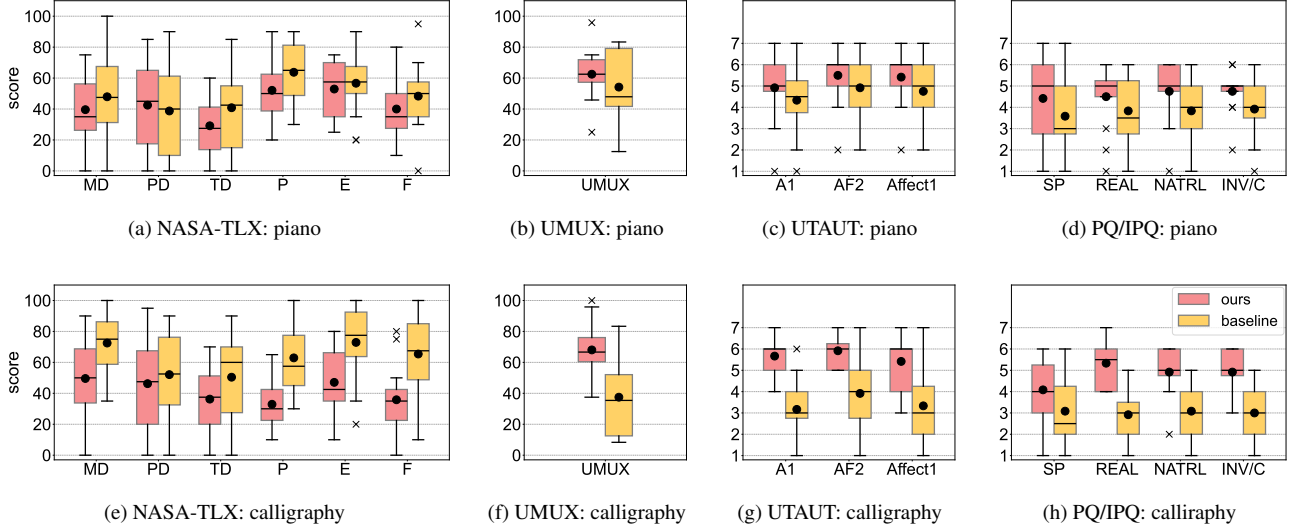


Figure 6: *Subjective evaluation of multimodal task learning.* Our post-study survey contains 17 questions (Table 3 in appendix) sampled from 5 standard XR/HCI questionnaires: NASA-TLX, UMUX, UTAUT, PQ, and IPQ. The Likert-scale scores are merged into 13 sub-groups to assess both conditions along 4 dimensions: subjective workload, perceived usability, technology acceptability, and sense of presence. *Higher* scores indicate better performance in UMUX, UTAUT, PQ, and IPQ, while *lower* scores are better in NASA-TLX. The boxes show the data range between the first and third quartiles (Q1-Q3), with dots indicating the mean and lines indicating the median. The error bars show the farthest data points lying within $1.5 \times \text{IQR}$ (inter-quartile range) from the boxes. Outliers beyond these limits are indicated as black crosses.

evaluate users' imitation performance. Among all the imitation attempts completed within a trial, only the last one was recorded and compared with the reference recording to compute the evaluation metric. Both ForceSense and the baseline method were utilized to run the same set of imitation tasks. Our experimental setup is visualized in Figure 4a. Note that the volume plot is shown here for illustration purposes only and was not available to the subjects.

Procedure. We adopted 3 single-keystroke piano recordings showing 3 different peak volume levels in this study. The study had $(\# \text{condition} = 2) \times (\# \text{reference} = 3) \times (\# \text{repetition} = 3) = 18$ trials. We conducted a short warm-up session of 4 minutes before running the study to let participants familiarize themselves with the task. To minimize order-induced performance bias, the condition order was counterbalanced across participants, and the piano recording order was randomized. In addition, we conducted a post-study survey to assess both conditions from four dimensions: subjective workload, perceived usability, technology acceptability, and sense of presence. The survey contained 17 questions (Table 3 in appendix) sampled from 5 standard XR/HCI questionnaires: NASA-TLX [15], UMUX [28], UTAUT [55], PQ [60], and IPQ [46]. The entire piano study, including hardware setups, pre-study instructions, warm-up trials, and post-study survey, took around 40 minutes per subject.

4.3.2 Visual-Tactile Task: Calligraphy Work Imitation

Conditions. Besides our force-aware interface ForceSense, we included a hand-tracking interaction method as the baseline for performance comparison. For ForceSense, we again employed a large table as the physical proxy and mapped a specific area on it to the virtual canvas. Subjects were instructed to slide their fingers on the table to generate handwriting strokes. The current location of their index finger was used to determine the stroke position. The decoded finger forces were utilized to adjust the stroke size. For the baseline condition, we defined a horizontal plane over the virtual canvas and mapped the penetration depth of the subject's index finger relative to that plane to varying force levels. The depth range [0, 9] in mm was linearly mapped to the force range [0, 13] in N.

Table 2: *One-way ANOVA on the interaction method.* In both the piano and calligraphy case studies, the interaction method exhibits a significant main effect on the users' task performance.

calligraphy work	$F_{1,11}$	p
1	24.6	< .0005
2	25.5	< .0005
3	27.3	< .0005
4	29.9	< .0005
5	13.7	< .005
6	16.8	< .005
overall	46.4	< .0001
peak volume	$F_{1,11}$	p
small	0.0152	ns
medium	10.0	< .01
large	2.06	> .05
overall	5.43	< .05

Tasks. We designed a visual imitation task to evaluate users' performance in reproducing several calligraphy works. Particularly, subjects were instructed to control the stroke size by adjusting the force intensity level. During each trial, a reference calligraphy work in light gray was shown on the canvas, and subjects were prompted to reproduce it using continuous handwriting strokes. A bright cursor indicating the projected location of the subject's index fingertip was rendered for visual guidance. Each trial had a fixed time limit of 30s. The intersection over union (IoU) metric between the reference and the imitated work was computed for performance evaluations. Both ForceSense and the baseline interaction method were utilized to reproduce the same set of calligraphy works. Figure 4b illustrates our experimental setup.

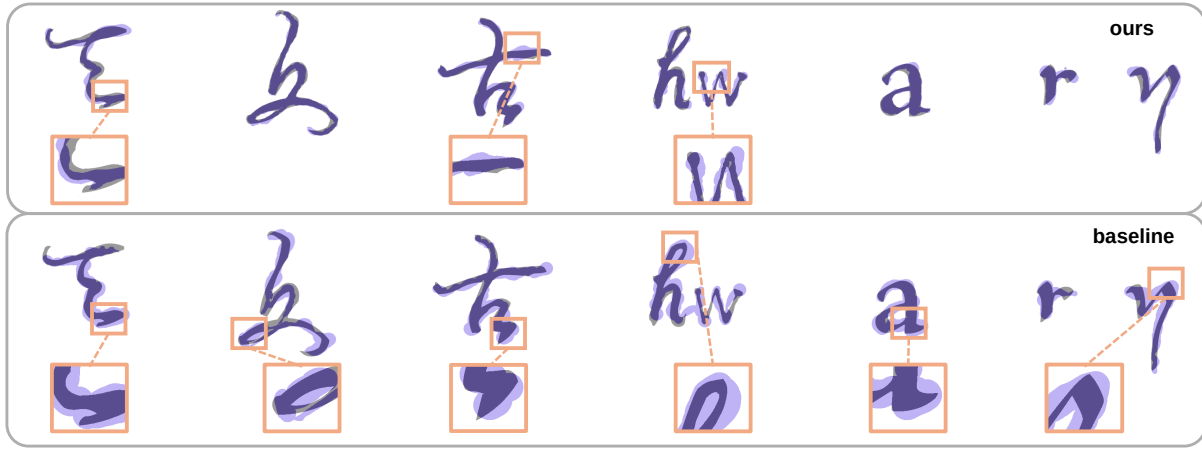


Figure 7: Randomly sampled imitation results from the calligraphy study. The reference calligraphy works (in gray) are overlaid with users’ imitation results (in purple). The top and bottom rows show a few samples obtained with *ForceSense* and the baseline method, respectively.

Procedure. We prepared 6 calligraphy works containing Chinese and English characters in this experiment. In total, there were $(\#condition = 2) \times (\#reference = 6) \times (\#repetition = 3) = 36$ trials. We conducted a short warm-up session of 4 minutes before running the study to let participants familiarize themselves with the task. To reduce order-induced bias, the condition order was counterbalanced across subjects, and the reference order was randomized. The same post-study survey from Section 4.3.1 was used. The complete study, including hardware setup, pre-study instructions, warm-ups, and post-study survey, took around 70 minutes per subject.

4.3.3 Results

Figures 5a and 5b illustrate the aggregated imitation errors of all 12 subjects for both studies. In the piano-playing study, the hand-tracking baseline is on par with *ForceSense* for piano recordings with small peak volume but largely falls behind for recordings with medium to large peak volume. Notably, *ForceSense* enables users to reproduce piano keystrokes with a mean volume error of 3.26dB across the entire volume range, which is 32.1% lower than the baseline. For the calligraphy-imitation study, *ForceSense* gives consistently better IoU scores across all 6 calligraphy works, and outperforms the baseline by 11.1% on average. Figure 7 illustrates a few randomly sampled imitation results. As shown in Table 2, one-way repeated measures ANOVA shows a significant main effect of the interaction method on subjects’ imitation performance in both studies. The Likert-scale scores aggregated from the post-study survey (Figure 6) validate that, compared with the hand-tracking interaction method, our force-aware interface *ForceSense* requires lower physical and mental workload to work with, presents better usability, and offers a stronger sense of presence to VR users.

4.3.4 Discussion

The results above demonstrate that *ForceSense* can effectively enhance VR users’ task learning performance in virtual environments. While the hand-tracking baseline and *ForceSense* both take a force-driven approach to implement hand-object interactions, *ForceSense* offers a more intuitive interface for VR users to interact with virtual objects. Importantly, *ForceSense* allows VR users to naturally exert physical forces without constraints on the hand and finger postures. The appropriate amount of forces are then decoded and applied to the object they intend to work with. Besides objective evaluations, the post-study questionnaire measuring users’ subjective preference also suggests that *ForceSense* is superior to the hand-tracking baseline in terms of workload requirements, technology acceptability, perceived usability, and sense of presence.

We argue that these benefits come from the fact that *ForceSense* enables a good synergy between visual, auditory, and haptic senses. In reality, humans naturally explore their surroundings by viewing, listening, and touching the objects therein. Faithfully recreating this multimodal interactive experience in virtuality is an indispensable component to designing user-friendly, successful VR interfaces.

5 LIMITATIONS AND FUTURE WORK

This research contributes a bare-hand force-aware VR interface and investigates its effectiveness in enhancing natural user embodiment in VR within both elemental interaction cases and practical application scenarios. Nonetheless, it also admits the following limitations for further research and future explorations.

First, a limited set of hand-object interactions (pulling, pushing, pressing) and their corresponding arm/hand postures were studied in this work. It is unclear whether the resulting *ForceSense* model would generalize to other unseen interaction types (such as squeezing and holding) under large variations of arm/hand postures.

Second, this research primarily studied force-driven interactions under moderate-to-high force regimes. For instance, the force range considered in the multimodal case studies of virtual calligraphy and piano playing (2-15N) was larger than that in the real world. Further experiments are necessary to validate the accuracy and robustness of *ForceSense* under lower finger force regimes.

Third, the current workflow requires a calibration session for new users to accommodate individual muscular variances. The associated preparation time, while not long, may compromise users’ perceived usability and thus lower their willingness to adopt the interface. In the future, we aim to explore online learning techniques [40] to achieve calibration-free deployment of *ForceSense*.

6 CONCLUSION

In this paper, we present *ForceSense*, an sEMG-to-force neural network model, and its corresponding adaptive interface for VR. With a series of objective measurements, human subject evaluation, and application case studies, we showed how narrowing the Real2Sim gap via a tactile-visual integrated interaction advances users’ real-world task efficiency and accuracy. We envision the research to provide the initial steps in the community to explore new opportunities in multi-sensory synergies for immersive interaction.

ACKNOWLEDGMENTS

This work is partially supported by NSF grant 2232817.

REFERENCES

- [1] Delsys trigno wireless surface electromyography sensing system, 2024. <https://delsys.com/>.
- [2] Sensel morph trackpad, 2024. <https://morph.sensel.com/>.
- [3] R. Albert, A. Patney, D. Luebke, and J. Kim. Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 14(4):1–13, 2017.
- [4] J. E. Aspell, L. Heydrich, G. Marillier, T. Lavanchy, B. Herbelin, and O. Blanke. Turning body and self inside out: visualized heartbeats alter bodily self-consciousness and tactile perception. *Psychological science*, 24(12):2445–2453, 2013.
- [5] B. Bardizbanian, J. Keating, X. Huang, and E. A. Clancy. Estimating individual and combined fingertip forces from forearm emg during constant-pose, force-varying tasks. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3134–3137. IEEE, 2020.
- [6] V. Becker, P. Oldrati, L. Barrios, and G. Sörös. Touchsense: classifying finger touches and measuring their force with an electromyography armband. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, pp. 1–8, 2018.
- [7] G. Caggianese, L. Gallo, and P. Neroni. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation. In *Intelligent Interactive Multimedia Systems and Services: Proceedings of 2018 Conference 11*, pp. 24–33. Springer, 2019.
- [8] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE transactions on neural systems and rehabilitation engineering*, 27(4):760–771, 2019.
- [9] P. Dourish. *Where the Action Is: The Foundations of Embodied Interaction*. MIT Press, Cambridge, MA, USA, 2001.
- [10] Y. Du, Y. Wong, W. Jin, W. Wei, Y. Hu, M. S. Kankanhalli, and W. Geng. Semi-supervised learning for surface emg-based gesture recognition. In *IJCAI*, pp. 1624–1630, 2017.
- [11] Y. Fang, D. Zhou, K. Li, Z. Ju, and H. Liu. Attribute-driven granular model for emg-based pinch and fingertip force grand recognition. *IEEE transactions on cybernetics*, 51(2):789–800, 2019.
- [12] M. Gazzoni, N. Celadon, D. Mastrapasqua, M. Paleari, V. Margaria, and P. Ariano. Quantifying forearm muscle activity during wrist and finger movements by means of multi-channel electromyography. *PloS one*, 9(10):e109943, 2014.
- [13] E. J. Gonzalez, E. Ofek, M. Gonzalez-Franco, and M. Sinclair. X-rings: A hand-mounted 360° shape display for grasping in virtual reality. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 732–742, 2021.
- [14] P. Gulati, Q. Hu, and S. F. Atashzar. Toward deep generalization of peripheral emg-based human-robot interfacing: A hybrid explainable solution for neurobotic systems. *IEEE Robotics and Automation Letters*, 6(2):2650–2657, 2021.
- [15] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.
- [16] R. Hu, X. Chen, H. Zhang, X. Zhang, and X. Chen. A novel myoelectric control scheme supporting synchronous gesture recognition and muscle force estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1127–1137, 2022.
- [17] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng. A novel attention-based hybrid cnn-rnn architecture for semg-based gesture recognition. *PloS one*, 13(10):e0206049, 2018.
- [18] S. Irawati, S. Green, M. Billingham, A. Duenser, and H. Ko. "move the couch where?": developing an augmented reality multimodal interface. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 183–186. IEEE, 2006.
- [19] J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. Trade-off between task accuracy, task completion time and naturalness for direct object manipulation in virtual reality. *Multimodal Technologies and Interaction*, 6(1):6, 2022.
- [20] I. Ketykó, F. Kovács, and K. Z. Varga. Domain adaptation for semg-based gesture recognition with recurrent neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. IEEE, 2019.
- [21] C. Khundam, V. Vorachart, P. Preeyawongsakul, W. Hosap, and F. Noël. A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training. In *Informatics*, vol. 8, p. 60. MDPI, 2021.
- [22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [23] N. Kondo, S. Kuroki, R. Hyakuta, Y. Matsuo, S. S. Gu, and Y. Ochiai. Immersive real world through deep billboards. In *ACM SIGGRAPH 2022 Immersive Pavilion*, pp. 1–2, 2022.
- [24] A. Kritikos and C. Brasch. Visual and tactile integration in action comprehension and execution. *Brain research*, 1242:73–86, 2008.
- [25] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE, 2010.
- [26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [27] S. Lee, J. Seo, G. J. Kim, and C.-M. Park. Evaluation of pointing techniques for ray casting selection in virtual environments. In *Third international conference on virtual reality and its application in industry*, vol. 4756, pp. 38–44. SPIE, 2003.
- [28] J. R. Lewis. Measuring perceived usability: The csuq, sus, and umux. *International Journal of Human-Computer Interaction*, 34(12):1148–1156, 2018.
- [29] P. Liu, F. Martel, D. Rancourt, E. A. Clancy, and D. R. Brown. Fingertip force estimation from forearm muscle electrical activity. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 2069–2073. IEEE, 2014.
- [30] Y. Liu, S. Zhang, and M. Gowda. Neuropose: 3d hand pose tracking using emg wearables. In *Proceedings of the Web Conference 2021*, pp. 1471–1482, 2021.
- [31] B. Long, S. A. Seah, T. Carter, and S. Subramanian. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Transactions on Graphics (TOG)*, 33(6):1–10, 2014.
- [32] P. Lopes, A. Ion, and P. Baudisch. Impacto: Simulating physical impact by combining tactile stimulation with electrical muscle stimulation. In *Proceedings of the 28th annual ACM symposium on user interface software & technology*, pp. 11–19, 2015.
- [33] P. Lopes, S. You, L.-P. Cheng, S. Marwecki, and P. Baudisch. Providing haptics to walls & heavy objects in virtual reality by means of electrical muscle stimulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 1471–1482, 2017.
- [34] Y. Luo, C. Liu, Y. J. Lee, J. DelPreto, K. Wu, M. Foshey, D. Rus, T. Palacios, Y. Li, A. Torralba, et al. Adaptive tactile interaction transfer via digitally embroidered smart gloves. *Nature Communications*, 15(1):868, 2024.
- [35] T. Luong, Y. F. Cheng, M. Möbus, A. Fender, and C. Holz. Controllers or bare hands? a controlled evaluation of input techniques on interaction performance and exertion in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2023.
- [36] H. Mao, P. Fang, and G. Li. Simultaneous estimation of multi-finger forces by surface electromyography and accelerometry signals. *Biomedical Signal Processing and Control*, 70:103005, 2021.
- [37] M. Meier, P. Strelai, A. Fender, and C. Holz. Tapid: Rapid touch interaction in virtual reality using wearable sensing. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 519–528. IEEE, 2021.
- [38] V. T. Nguyen and T. Dang. Setting up virtual reality and augmented reality learning environment in unity. In *2017 IEEE International symposium on mixed and augmented reality (ISMAR-Adjunct)*, pp. 315–320. IEEE, 2017.
- [39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8026–8037, 2019.
- [40] B. Pérez-Sánchez, O. Fontenla-Romero, and B. Guijarro-Berdiñas. A review of adaptive online learning for artificial neural networks. *Arti-*

ficial Intelligence Review, 49:281–299, 2018.

- [41] A. Phinyomark, E. Campbell, and E. Scheme. Surface electromyography (emg) signal processing, classification, and practical considerations. *Biomedical Signal Processing: Advances in Theory, Algorithms and Applications*, pp. 3–29, 2020.
- [42] A. Phinyomark and E. Scheme. Emg pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing*, 2(3):21, 2018.
- [43] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billingham. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE symposium on 3D user interfaces (3DUI)*, pp. 36–39. IEEE, 2017.
- [44] B. Rodríguez-Tapia, I. Soto, D. M. Martínez, and N. C. Arballo. Myoelectric interfaces and related applications: current state of emg signal processing—a systematic review. *IEEE Access*, 8:7792–7805, 2020.
- [45] D. Sandykbayeva, Z. Kappasov, and B. Orazbayev. Vibrotouch: Active tactile sensor for contact detection and force sensing via vibrations. *Sensors*, 22(17):6456, 2022.
- [46] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001.
- [47] J.-K. Song, D. Son, J. Kim, Y. J. Yoo, G. J. Lee, L. Wang, M. K. Choi, J. Yang, M. Lee, K. Do, et al. Wearable force touch sensor array using a flexible and transparent electrode. *Advanced Functional Materials*, 27(6):1605286, 2017.
- [48] P. Strelci, J. Jiang, A. R. Fender, M. Meier, H. Romat, and C. Holz. Taptype: Ten-finger text entry on everyday surfaces via bayesian inference. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2022.
- [49] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*, 569(7758):698–702, 2019.
- [50] Y. Tanaka, A. Shen, A. Kong, and P. Lopes. Full-hand electro-tactile feedback without obstructing palmar side of hand. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2023.
- [51] A. Toro-Ossaba, J. Jaramillo-Tigeros, J. C. Tejada, A. Peña, A. López-González, and R. A. Castanho. Lstm recurrent neural network for hand gesture recognition using emg signals. *Applied Sciences*, 12(19):9700, 2022.
- [52] P. M. Torrens and R. Kim. Evoking embodiment in immersive geosimulation environments. *Annals of GIS*, 30(1):35–66, 2024. doi: 10.1080/19475683.2024.2316601. doi: 10.1080/19475683.2024.2316601
- [53] D. Valkov and L. Linsen. Vibro-tactile feedback for real-world awareness in immersive virtual environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 340–349. IEEE, 2019.
- [54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [55] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS quarterly*, pp. 425–478, 2003.
- [56] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 27–35, 2016.
- [57] D. Wang, K. Ohnishi, and W. Xu. Multimodal haptic display for virtual reality: A survey. *IEEE Transactions on Industrial Electronics*, 67(1):610–623, 2019.
- [58] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H.-N. Liang. Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2478–2488, 2023.
- [59] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanalli, and W. Geng. Surface-electromyography-based gesture recognition by multi-view deep learning. *IEEE Transactions on Biomedical Engineering*, 66(10):2964–2973, 2019.
- [60] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.
- [61] J. Xue and K. W. C. Lai. Dynamic gripping force estimation and reconstruction in emg-based human-machine interaction. *Biomedical Signal Processing and Control*, 80:104216, 2023.
- [62] L. Yang, J. Huang, T. Feng, W. Hong-An, and D. Guo-Zhong. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(1):84–112, 2019.
- [63] Y. Yeshurun, M. Carrasco, and L. T. Maloney. Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision research*, 48(17):1837–1851, 2008.
- [64] X. Yu, Z. Xie, Y. Yu, J. Lee, A. Vazquez-Guardado, H. Luan, J. Ruban, X. Ning, A. Akhtar, D. Li, et al. Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature*, 575(7783):473–479, 2019.
- [65] X. Zhai, B. Jelfs, R. H. Chan, and C. Tin. Self-recalibrating surface emg pattern recognition for neuroprosthesis control based on convolutional neural network. *Frontiers in neuroscience*, 11:266372, 2017.
- [66] Y. Zhang, B. Liang, B. Chen, P. M. Torrens, S. F. Atashzar, D. Lin, and Q. Sun. Force-aware interface via electromyography for natural vr/ar interaction. *ACM Transactions on Graphics (TOG)*, 41(6):1–18, 2022.
- [67] Y. Zhou, C. Chen, M. Cheng, Y. Alshahrani, S. Franovic, E. Lau, G. Xu, G. Ni, J. M. Cavanaugh, S. Muh, et al. Comparison of machine learning methods in semg signal processing for shoulder motion recognition. *Biomedical Signal Processing and Control*, 68:102577, 2021.
- [68] M. Zhu, Z. Sun, Z. Zhang, Q. Shi, T. He, H. Liu, T. Chen, and C. Lee. Haptic-feedback smart glove as a creative human-machine interface (hmi) for virtual/augmented reality applications. *Science Advances*, 6(19):eaaz8693, 2020.
- [69] T. Zindulka, M. Bachynskyi, and J. Müller. Performance and experience of throwing in virtual reality. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–8, 2020.