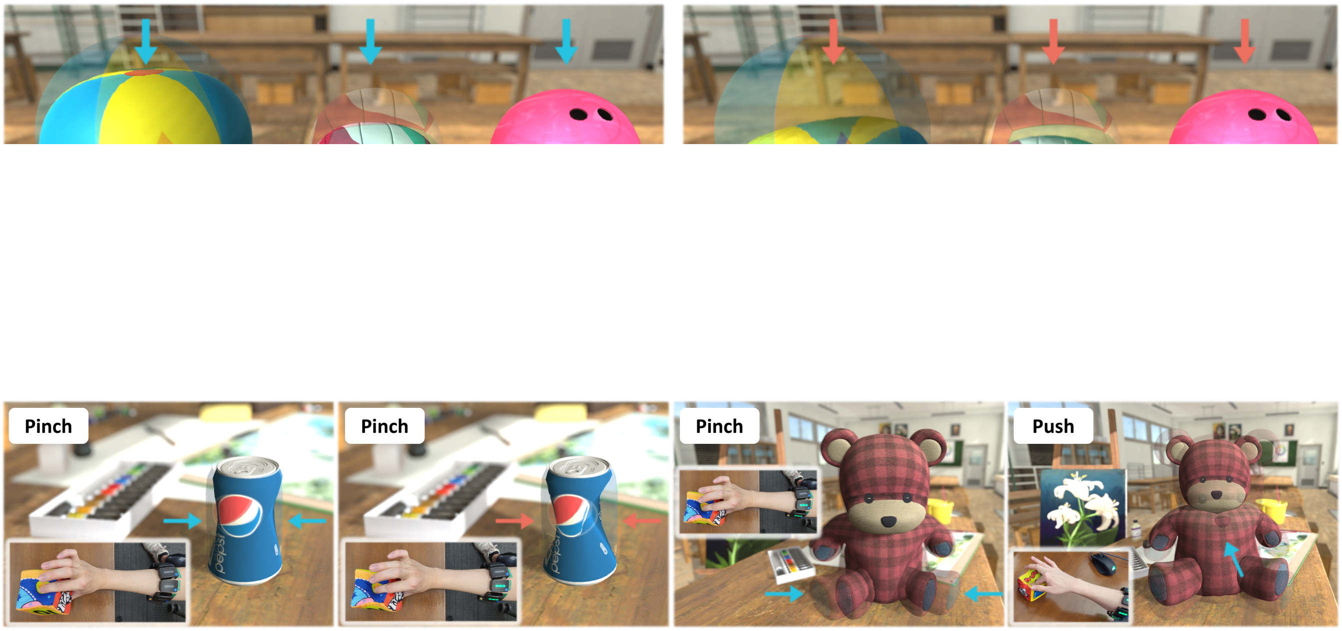
Force-Aware Interface via Electromyography for Natural VR/AR Interaction

YUNXIANG ZHANG, New York University, USA and The Chinese University of Hong Kong, Hong Kong SAR



(b) Our system provides a natural and intuitive interface for capturing user-generated forces and letting them take effects in the virtual environment.

Fig. 1. *Electromyography (EMG)-based neural interface with learned muscular force decoder.* With user-generated physical forces being decoded by our interface and applied to virtual objects in real-time, (a) illustrates the deformation of a beach ball, a volleyball, and a bowling ball in VR, subject to pressing force of varying intensities. The beach ball is softer than the volleyball and thus exhibits larger deformation under the same force level, while the bowling ball is rigid and barely deforms within the force range of finger pressing. Our scheme helps users better perceive/distinguish the physical properties of virtual objects, in a way similar to how they approach it in the real world. (b) shows force-enabled virtual interactions via our system, with enhanced physical realism. 3D asset credits to SbbUtutuya, Virtual Method, TGameAssets at Unity, and TankStorm at Sketchfab.

While tremendous advances in visual and auditory realism have been made for virtual and augmented reality (VR/AR), introducing a plausible sense of physicality into the virtual world remains challenging. Closing the gap between real-world physicality and immersive virtual experience requires a closed interaction loop: applying user-exerted physical forces to the virtual environment and generating haptic sensations back to the users. However, existing VR/AR solutions either completely ignore the force inputs from the users or rely on obtrusive sensing devices that compromise user experience.

By identifying users' muscle activation patterns while engaging in VR/AR, we design a learning-based neural interface for natural and intuitive force inputs. Specifically, we show that lightweight electromyography sensors, resting non-invasively on users' forearm skin, inform and establish a robust understanding of their complex hand activities. Fuelled by a neural-network-based model, our interface can decode finger-wise forces in real-time with 3.3% mean error, and generalize to new users with little calibration. Through

Authors' addresses: Yunxiang Zhang, yunxiang.zhang@nyu.edu, New York University, USA and The Chinese University of Hong Kong, Hong Kong SAR; Benjamin Liang, ben.liang@nyu.edu; Boyuan Chen, boyuan.chen@nyu.edu, New York University, USA; Paul M. Torrens, pt50@nyu.edu; S. Farokh Atashzar, sfa7@nyu.edu, New York University, USA; Dahua Lin, dhlin@ie.cuhk.edu.hk, The Chinese University of Hong Kong, Hong Kong SAR and Shanghai Artificial Intelligence Laboratory, China; Qi Sun, qisun@nyu.edu, New York University, USA.

0730-0301/2022/12-ART268 $15.00
https://doi.org/10.1145/3550454.3555461

an interactive psychophysical study, we show that human perception of virtual objects' physical properties, such as stiffness, can be significantly enhanced by our interface. We further demonstrate that our interface enables ubiquitous control via finger tapping. Ultimately, we envision our findings to push forward research towards more realistic physicality in future VR/AR.

CCS Concepts: • **Computing methodologies → Virtual reality**; **Mixed / augmented reality**; **Perception**; **Neural networks**; • **Human-centered computing → Haptic devices**.

Additional Key Words and Phrases: Electromyography, Force-Aware Neural Interface, Machine Learning, Haptic Perception

ACM Reference Format:
Yunxiang Zhang, Benjamin Liang, Boyuan Chen, Paul M. Torrens, S. Farokh Atashzar, Dahua Lin, and Qi Sun. 2022. Force-Aware Interface via Electromyography for Natural VR/AR Interaction. *ACM Trans. Graph.* 41, 6, Article 268 (December 2022), 18 pages. https://doi.org/10.1145/3550454.3555461

1 INTRODUCTION

The *visual* gaps between real world and virtual environments have been rapidly shrinking with the advance of novel display and rendering technologies. However, developing matching realistic-*feeling* interfaces that let users interact as if they were in the physical world, stands out as a chronically persistent and doggedly resistant challenge [Torrens and Gu 2021]. Physical interactions, such as lifting, grasping, brushing, pushing, and prodding involve a bi-directional interchange between humans and the environment: our muscles exert forces on objects, while we perceive the visual (and sometimes haptic) feedback response in the objects' reactions. To establish the same loop in virtual environments, researchers have devoted extensive effort to advance the quality of feedback sensations with haptic devices and rendering methods. However, it has long remained difficult to transfer real-world physical human applied forces of dexterity and agility into convincing virtual form. This incomplete loop leaves VR/AR disadvantaged in its ability to faithfully and convincingly represent real experiences.

The idea of directly sensing, tracking, and decoding user-induced forces has emerged as a promising line of research, with the implication that this information provides a scaffold for building natural and intuitive interaction experiences [Bergström and Hornbæk 2019; Ernst and Banks 2002]. Wearable force sensors now provide high-precision and high-resolution data [Luo et al. 2021a; Sundaram et al. 2019]. However, existing force-sensing technologies are often bulky, wired, and directly attached to hands. This hampers their applications for natural interaction and makes the devices undesirable as consumer-level interfaces. One solution has been to skip devices altogether. Purely data-driven visual-to-force learning methods [Ehsani et al. 2020a] have been proposed, allowing for contactless *estimation* of user force. However, they may suffer from occlusions, low precision, and action-perception delays due to the high load of transmission and processing.

The advancement of neural sensing enabled central (from the brain [Anumanchipalli et al. 2019; Willett et al. 2021]) and peripheral (from the muscles [Liu et al. 2021; Salemi Parizi et al. 2021]) solutions for decoding human action intentions, as electrophysiological responses. However, decoding the intended force for interaction has been so far unsolved due to the variance across human users, the lack of correlated data, computational complexity that blocks real-time performance, and the inevitable pervasive sensory noises affecting biological signals [Hof 1991].

We introduce an end-to-end neural interface that reduces the physicality gap between real experience and VR/AR. The result is a real-time system for *dexterity-enabled force-aware VR/AR*. The system's chief advantages are that, (1) it allows for natural, unimpeded, forearm and hand movement; (2) using off-the-shelf electromyography sensors; (3) with low latency for force-and-response interactions with computer graphics; (4) in ways that are generalizable across a diversity of users.

Our research shows that very detailed and rich physical experiences of manual dexterity can be delivered to VR/AR systems and paired with high-fidelity graphics for visual similitude, in ways that neatly and realistically close the loop between intention and interaction in VR/AR experiences. Our system offers a tractable solution to existing bottlenecks in directly sensing and resolving users' physical intentions in VR/AR systems, using machine-learning on skin-surface electromyography (sEMG) sensors to identify, track, and decode signals of physical activity at rates that allow for matching design and delivery of experiential content in VR/AR. These developments, while preliminary, open-up new pathways for VR/AR experience in gaming, design, and object control. While we describe the research, development, and evaluation pipeline, we note that the system is application-ready. We demonstrate practical examples on low-cost commercially available sEMG sensors and widely used VR/AR technology. We will also open-source both our dataset and source code to the community to support future work.

While the system is shown to work parsimoniously in user-testing and evaluation, the research behind it is non-trivial. To develop the proposed neural interface, we start by collecting a large-scale joint dataset via force-sensing and surface EMG devices. The dataset consists of the time-synchronized signals between fingertip forces and the corresponding EMG signals. By leveraging our specialized dataset, we developed the first real-time learning-based framework that tracks and decodes human physical forces from multichannel muscle activation signals. The dataset is populated, initially, using a set of participant experiments to record EMG signals from forearm muscles while participants directly perform various natural hand-object interactions, such as pressing and pinching. On this initial dataset, we trained a convolutional neural network (CNN) model on the frequency-transformed signals to robustly learn the complex mapping between muscle activities and actions. The trained model isolates the force-induced bio-electrical signals from hand motions and estimates the forces exerted at the fingertips. During run-time, the model only uses the *past* 624ms of EMG data, enabling low-latency force inference in real-time. With this model on hand, we show that only minimal calibration is required to transfer and generalize it to unseen users.

In order to validate the system we have conducted a systematic user study and evaluated users' experience during interaction with virtual objects. We will present a series of psychophysical experiments and objective analysis that reveal our system to be robust and generalizable. Moreover, we will show that the system can enhance users' perceptual understanding of virtual objects' physical and material characteristics in VR, by extending their capabilities for natural human interaction with graphical objects. Our experimentation also demonstrates that the system is broadly resilient to variation in user physiology, sensor placement, and tasks.

In summary, this paper contributes:

- An end-to-end EMG-based neural interface that decodes, transfers, and applies hand-induced forces with low-latency in VR environments;
- A prototype interaction system that leverages our method to enhance human's perceptual understanding of material characteristics in VR;
- A real-time and generalizable CNN-based model established in the frequency demain of EMG-sensed muscular potentials with a force-tailored loss design;
- A set of user experiments to demonstrate the generalizability of the system to perturbations in sensor placement, shifting task context, and uniqueness of users;
- Proofs of concept for natural interaction with computer graphics in VR.

We provide the source code for our force regression models, real-time interaction system, and accompanying EMG-Force dataset at https://github.com/NYU-ICL/xr-emg-force-interface.

2 RELATED WORK

2.1 Biometric Sensing for Immersive Interaction

Accurately sensing human behaviors is fundamental to favorable human-environment interaction. With recent advancements of various sensing technologies, both in hardware and software, we are now entering an era where unprecedented means of multi-modal interaction with virtual environments are possible. For instance, eye tracking enables real-time foveated rendering [Kim et al. 2019; Patney et al. 2016] and enhances VR redirected walking [Langbehn et al. 2018; Sun et al. 2018]; face tracking generates lifelike virtual avatars for telecommunication [Chen et al. 2021; Chu et al. 2020; Ma et al. 2021]; whole-body tracking allows for intuitive control and feedback for virtual interaction [Cao et al. 2017; Joo et al. 2018; Kanazawa et al. 2018; Newell et al. 2016]. In virtual environments, users largely rely on hand-based interfaces for interaction, making hand behaviors particularly indicative of their intention and status. As a result, hand tracking has attracted considerable research interest in computer graphics [Boukhayma et al. 2019; Han et al. 2020; Romero et al. 2017; Wan et al. 2018]. However, tracked position information alone is insufficient to achieve immersive VR experience. Without the *feeling* of hands, positional tracking essentially casts ghost appendages in users' field of view. This misses the sense of *corporeality* and thus the sense of capabilities that humans feel as they use their hands in the real world. We reason that *hand-induced interaction force* is another indispensable component of human embodiment in virtual scenes that is often overlooked or only approximated [Pham et al. 2015; Zhu et al. 2016] in prior works.

2.2 Sensing and Interacting with Contact Forces

Contact forces are an essential modality for understanding and enhancing human-object interaction [Luo et al. 2021b; Sundaram et al. 2019]. Unlike visual stimuli, force information must be communicated in a *two-way* fashion when we interact with and establish understanding of virtual environments. While users apply forces to a virtual object, they also receive haptic feedback from the object's response [Dangxiao et al. 2019; Gonzalez et al. 2021; Yoshida et al. 2020]. For the latter, which has been addressed in computer graphics as haptic rendering [Lin and Otaduy 2008], researchers have explored various ways of applying tactile effects to users, ranging from grasping and touching [Choi et al. 2016, 2018; Verschoor et al. 2020] to texture [Benko et al. 2016], shear [Whitmire et al. 2018] and gravity [Choi et al. 2017].

However, the inverse problem of naturally sensing and exploiting human-exerted forces in the context of VR remains an open challenge. Existing solutions are either based on hand-held input devices or force-sensing wearables [Luo et al. 2021a; Sundaram et al. 2019]. While such methods can provide high-precision force measurements during hand-object interaction, their *obtrusive* design inevitably compromises finger dexterity, increases the *frictions* between users and virtual environments, and limits their availability for daily usage. To develop a natural and intuitive force interface for VR, we attempt to sense hand-applied forces from the controlling muscles located on the forearm by leveraging the biological mechanism of human hands as described in Section 3.1. This allows us to completely bypass on-hand measurements and achieve force-enabled VR interaction in a natural bare-hand manner.

2.3 EMG-Based Human-Computer Interface

Recent advancements in neural interfaces have demonstrated the great potential of interactive devices that directly interface with the human body and interpret neuronal activities for downstream tasks [Anumanchipalli et al. 2019; Flesher et al. 2021; Hochberg et al. 2012; Willett et al. 2021]. Among these interfaces, EMG has emerged as a promising interaction medium, especially in VR and AR [Hirota et al. 2018; Koniaris et al. 2016; Tsuboi et al. 2017]. A major benefit of EMG for immersive interaction is the potential that it offers for completely bypassing the often-used solution of camera-based tracking, which has serious side effects of being open to limitation by occlusions and field of view [Pai et al. 2019]. To advance EMG approaches, considerable research efforts have been made to infer hand poses from forearm EMG, including gesture recognition [Du et al. 2017; Gulati et al. 2021; Javaid et al. 2021; Jo and Oh 2020; Rahimian et al. 2021; Sun et al. 2022], hand orientation estimation [Andrean et al. 2019; Zhao et al. 2020], and finger tracking [Liu et al. 2021; Qi et al. 2021; Zhang et al. 2022]. The knowledge may then be leveraged towards camera-free VR control [Ahsan et al. 2009]. We reason, additionally, that tracking hand-object *interaction forces* is indispensable to creating realistic physical effects in VR, e.g., via physics-based simulation methods. However, finger-exerted forces are continuous, transient, subtle, and changeable, thus pitching fundamental challenges for decoding.

Prior research investigated the possibility of estimating hand/finger-level forces from forearm EMG [Baldacchino et al. 2018; Bardizbanian et al. 2020a,b; Becker et al. 2018; Castellini and Koiva 2012; Castellini and Van Der Smagt 2009; Cho et al. 2022; Fang et al. 2019; Gailey et al. 2017; Hu et al. 2022; Liu et al. 2013; Mao et al. 2021; Martínez et al. 2020; Martinez et al. 2020; Wu et al. 2020, 2021; Zhang et al. 2022]. Despite exciting preliminary results, deploying them in practical VR applications is still in its infancy. Several open problems remain mostly unresolved:

Flexibility for real-life usage. Existing solutions commonly assume controlled laboratory settings. For example, work presented by Castellini and Koiva [Castellini and Koiva 2012] can only operate when the user's hand is artificially pinned and constrained in a flat wooden mold/guide. In the approach by Zhang et al. [Zhang et al. 2022], it is necessary to place hard-wired electrodes up and down an entire arm; moreover, force detection relies on an elaborate mechanical metal force-sensing device that is hard-bolted to a table. In the approach shown by Baldacchino et al. [Baldacchino et al. 2018], there is also a requirement that an entire arm be fitted with electrodes. These systems are fantastic early proofs-of-concept, but interacting in VR demands requires that free-form interaction is supported—we would argue that it also needs to be as natural as possible—and this necessitates a different approach over the current state-of-the-art to successfully bypass those complications. Our approach introduces a relaxed, accessible, natural test-bed that can accommodate freely realistic postures and gestures of the hand. The level of authenticity that we have achieved relative to real-world hand and finger forces contrasts with much of the prior art. Existing approaches are highly isometric, which artificially limits free interaction in testing and in use.

Simultaneous and continuous multi-finger force measurement. To reproduce natural dexterity, it is necessary to enable all fingers to operate together and apply varied forces simultaneously. This is critical to how we use our upper limbs to manipulate and explore the world around us. Most existing approaches to reproducing this in VR have focused on generalized hand-scale gestures [Fang et al. 2019; Gailey et al. 2017; Hu et al. 2022; Martínez et al. 2020; Wu et al. 2021]. Humans rely on most muscles in the forearm and neural control of these muscles produces electrical signals. These signals are notably unambiguous and thus open to direct detection. This is not always straightforward outside of clinical sensing. Beyond a classification problem, regressing the exact force value positions presents additional challenges due signal noise and individual variances. This has been tried before. For example, Baldacchino et al. [Baldacchino et al. 2018] presented a regression approach, but rather than sensing they tackled the challenge through data science on an existing database [Atzori and Müller 2015], which was limited to nine variations of a simple (and *single*) finger-on-surface pressing motion (compare this to the free-form and multi-finger dexterous gestures that our scheme tackles). Real humans of course use their fingers as they please during dexterous tasks; limiting dexterity to a single finger would seriously hamper usability. Continuity in the temporal domain presents another challenge. As one can imagine, the signal firing of muscles in the forearm is highly dynamic. Previous work have addressed this by approximating dynamics as "action shifts" between hand postures [Gailey et al. 2017]. This is really just a workaround that substitutes state transition for actual dynamics. This is problematic for VR settings, which are often highly dynamic, with users that are usually quite aware of how fast their hands and fingers move in the real world. Building realistic and fast-adaptive temporal continuity between user actions and responding force-aware graphics is therefore critical in supporting natural interaction.

Our approach, by comparison, *simultaneously* isolates and teases out signal details for *individual fingers* with *spatio-temporally continuous* force prediction. Our aim, in doing so, is to support a wide range of natural interactions in VR/AR, including those that require fine-grain dexterity maneuvers. The force-based dexterous abilities that are accessible via our scheme (e.g., finger tapping) are well beyond the capabilities of existing prior art (which focus on finger posture (not force) or track very stylized dexterity such as simple pressing actions). This is achieved via our machine learning approach and an in-house dataset with robust interaction variety.

Generalizability. Daily interactive scenarios require generalizable systems for everyone, without tedious pre-usage preparation. The prior art in this domain adopts an approach that validates cross-user force prediction accuracy with datasets under *identical settings* [Bardizbanian et al. 2020a,b; Becker et al. 2018; Castellini and Van Der Smagt 2009; Mao et al. 2021; Zhang et al. 2022]. By contrast, our *frequency domain neural network method* tackles the long-looming generalizability problem in EMG data decoding. In this paper, we show that we can collect data on two completely different days (with associated shifts in placement of sensors) as well as for completely different users (with shifting reactions, varying arm and finger morphology, and different dexterity and skills) with less than 2 minute calibration. Solving sensing generalizability and subject generalizability—in tandem—is a significant contribution to the literature. Moreover, it greatly expands the applicability of our scheme for VR/AR, where there will necessarily be wide variation gaps in sensing conditions and users.

3 METHOD

In the following, we first review the biological mechanism of human hands and illustrate how muscles on the forearm control finger-level forces in Section 3.1. Then, we describe how we tailored our EMG-force joint data collection to capture the complex mapping from muscle activations to finger-level forces in Section 3.2. Finally, using the established dataset, we detail our frequency-domain muscular force learning pipeline, with joint classification and regression, in Section 3.3.

3.1 Biological Model of Human Hands

To bypass the limitations of passively measuring hand-exerted interaction forces using cumbersome (and interfering) sensors such as gloves, we argue that such information can be actively decoded at the finger level from the bioelectric signals reflecting forearm muscle activations, which wireless EMG sensors can in turn capture.

Hand-forearm joint biomechanical mechanism. As shown in Figure 2 (bottom), hands, the most dexterous limbs on the human body, exhibit high degree-of-freedom (DOF) articulations through a large number of finger joints, allowing us to perform complex and subtle interactions with the surroundings. The muscles driving this delicate articulated structure are: 1) extrinsic muscles spread over the anterior and posterior compartments of the forearm; 2) intrinsic muscles located right in the hand.

Exten:
radiali

Fig. :
exte
at fi
pres

Fi
inte:
ing, pinching, and gripping [Von Hardenberg and Bérard 2001]. The major contributing muscles in these interactions include flexor digitorum superficialis, flexor digitorum profundus, and flexor pollicis longus, all residing in the anterior compartment of the forearm. In particular, flexor digitorum superficialis controls the flexion of PIP and MCP joints for the 4 fingers and the wrist; flexor digitorum profundus controls the flexion of DIP joints for the 4 fingers as well as the flexion of MCP joints for the 4 fingers and the wrist; flexor pollicis longus controls the flexion of IP and MCP joints for the thumb. Figure 2 illustrates the anatomical structure of these forearm flexor muscles. By investigating the signals passed when invoking these biomechanics, we reason that it becomes possible to *sense and learn* hand operations from the connected forearm.

Bioelectric mechanism. Muscles are composed of constituent elements called motor units, and the contraction of each single muscle is managed by a specific group of motor units. On the other hand, motor units are made from more fundamental units called muscle fibers. When activated by our brain, muscle fibers within the same motor unit fire together and generate a propagating electrical potential called motor unit action potential (MUAP) via the elevation of Ca^{2+} in the sarcoplasm [Melzer et al. 1984]. When placed on our skin, the EMG sensors record such electrical signals in real-time. The various hand-object interactions that we can perform are results of varying activation patterns of the involved muscle fibers, which are themselves reflected by the recorded signals. However, how analytically or numerically the electrical signals are coupled with mechanical forces remains an open challenge, especially given the inevitable sensing noise. In the following section, we discuss our attempts toward a robust electric-mechanic signal decoding in the frequency domain.

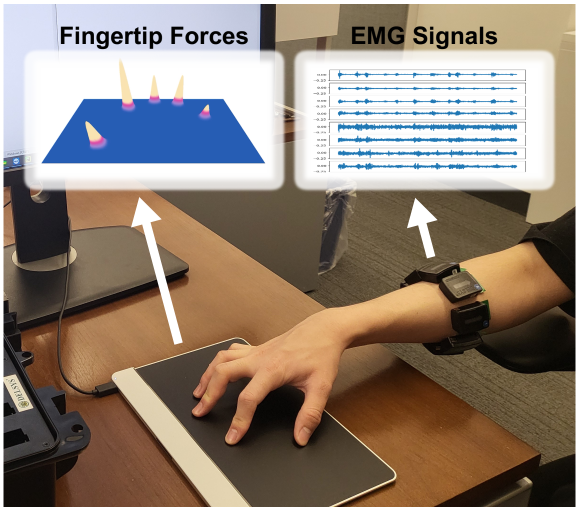


Fig. 3. *Experimental setup for collecting time-synchronized EMG and force data.* During the data collection, participants were asked to interact with a Morph Sensel trackpad through various pressing and pinching actions while wearing 8 EMG sensors on his/her right forearm.

3.2 EMG-Force Joint Data Collection

We aim to establish a bioelectrical-mechanical bridge via a data-driven approach. To this end, we first collect time-synchronized EMG and force data in a supervised manner. To collect EMG electrical signals, we adopt 8 Delsys Trigno EMG sensors (from Delsys Inc, USA) and overlay them on the forearm in a way that all muscles of interest are monitored. All 8 EMG sensors are wirelessly synchronized at 2000 Hz. We note that this is a factor of *ten times* the bandwidth of the (now discontinued) Myo sensor that is used in prior art, e.g., Javaid et al. [Javaid et al. 2021]. In their review of the accuracy of sEMG sensors, Pizzolato et al. [Pizzolato et al. 2017] discuss this issue directly, noting that “the Myo is not suited to record high quality sEMG signal data including the full power spectrum of sEMG (that can include frequencies of up to 300-500 Hz)” (p.10). Our captured data are streamed to a desktop computer over WiFi in real-time. To collect finger-wise force data, we employed a Morph Sensel trackpad with pressure sensors arranged into a dense array. The data collection setup is illustrated in Figure 3. In particular, we divide the tracking area into 5 non-overlapping regions so that each fingertip only taps onto its dedicated partition throughout the data collection process. Detected contact points with force information can then be correctly attributed to the corresponding fingers. Note that this design choice is only adopted to ease force labeling efforts and we do not assume any specific wrist/finger poses during either training or testing. Also, the dividing strategy is user- and motion-specific to accommodate the hand size and personal habit of different users. The data collection code for both modalities is launched using multi-threading, and the system timestamps are exploited for overall synchronization.

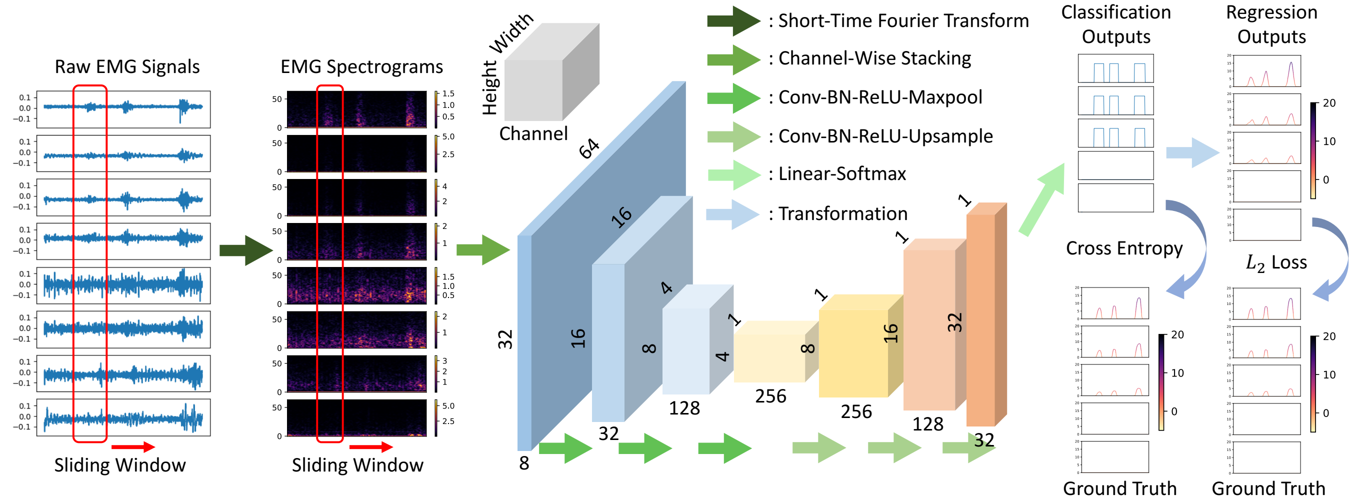


Fig. 4. *Illustration of the deep learning pipeline embedded in our system.* To optimize for parameter efficiency and resilience to data noise, we transform raw EMG signals into frequency domain via STFT, treat the resulting spectrograms as multi-channel images, and employ a lightweight CNN model with bottleneck design. Training is performed using a customized classification-regression joint loss tailored to the task of force estimation. When put into action, our system uses a fixed-size sliding window to retrieve the latest frames from wirelessly streamed EMG signals and decodes finger-wise forces in real time.

Transformed data representation. As a typical bio-electric sensor, EMGs also suffer from a certain level of measurement noise including powerline noise and other electromagnetic artifacts. Existing EMG processing approaches typically extract and learn from statistical features in the time domain, such as mean absolute value, average amplitude change, interquartile range, etc [Spiewak et al. 2018]. Consequently, subtle noise or distortion may cause significant feature-space error [Boostani and Moradi 2003], harming the change-sensitive force-bioelectricity correlation.

Drawing inspirations from audio research, we compute the spectrograms of EMG signals using short-time Fourier transform (STFT) so that high-frequency additive noise may be more distinctly isolated. Another computational advantage of learning with frequency-domain representation is that the EMG signal from each electrode, or channel, is now a 2D array instead of a 1D time series and that we can seamlessly take advantage of powerful convolutional neural network (CNN) models for better parameter efficiency and generalization capability. Specifically, we adopt a Hanning window of size 256 sample points, which corresponds to a duration of 128ms, with hop length set to 32, to obtain 129 frequency bins. In addition, a resampling step is needed to temporally align raw force data (the sampling frequency of Morph Sensel is around 125Hz) with computed EMG spectrograms. A nearest-neighbor-based interpolation is adopted for this purpose.

3.3 Muscular Force Learning Pipeline

A main roadblock for EMG sensors is the well-known challenge of aligning the electrodes exactly on muscles. For instance, as seen in Figure 2, sensors may commonly cross-ride on or fall in the gap between the underlying interwoven muscle bundles. As a result, although the activation information of all target muscles are captured by EMG sensors, directly assigning the electric signals to individual muscle-group and joints becomes unrealistic. To robustly recover finger-wise force information from raw EMG data, we resort to the data-driven paradigm and adopt powerful neural network models to learn this highly non-linear correlation between forearm EMG signal and finger-wise forces.

Model architecture. While recurrent neural network (RNN) has been a common practice for sequential data learning, recent advancements in audio learning have shown that deep CNN models with properly processed input data are capable of delivering better performance in some cases, thanks to their highly efficient parameter usage which allows for very deep design [Oord et al. 2016]. We are inspired to exploit convolutional filters to extract deep features from the 2D spectrograms. While our input data points live in a high-dimensional space (129), those features containing the semantic information of finger-wise forces are embedded in a subspace of much lower dimension. To efficiently extract relevant information and mitigate overfitting to training data, we employ an encoder-decoder architecture to enforce a low-dimensional latent space. Also, we only feed the 64 low-frequency components from the spectrograms to the encoder model to remove high-frequency data noise and allow for accelerated performance. In addition, we provide the model with sequential data from a long time interval (32 consecutive frames in the spectrograms, which correspond to 624ms raw EMG data) instead of a single frame to let it exploit information from previous frames and better satisfy temporal constraints. The input data size is thus (N, C, L, S), where N, $C = 8$, $L = 32$, and $S = 64$ denote the batch size, number of sEMG channels, input sequence length, and number of EMG frequency components, respectively. The encoder model consists of repeating Convolution-BatchNorm-ReLU blocks, with each block followed by a 2×2 Maxpool layer for downsampling in time and frequency dimensions. Similarly, the

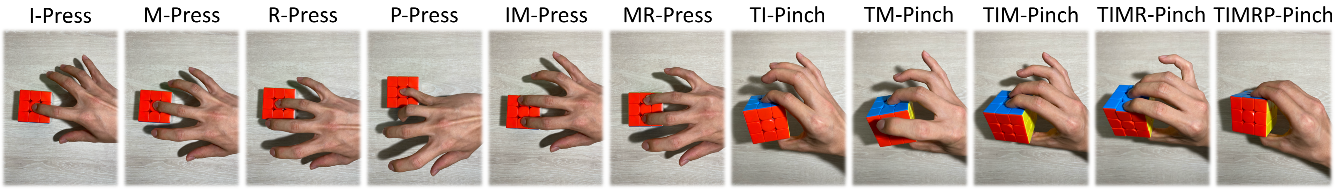


Fig. 5. *Types of hand-object interaction selected for constructing our EMG-Force dataset.* Capital letters before the hyphen, namely T, I, M, R, and P, stand for thumb, index finger, middle finger, ring finger, and pinky finger, respectively.

decoder model also consists of repeating Convolution-BatchNorm-ReLU blocks, with each block followed by a 2 × 2 bilinear Upsample layer for increasing time dimension. After that, the decoder output is transformed by a linear layer in the channel dimension to match desired force outputs, e.g. 5 values for 5 finger-wise forces. The output data size is thus $(N, F, L, 1)$, or (N, F, L) with the fake frequency dimension squeezed out, where F denotes the pre-defined number of force components. All convolutional layers have kernel size 3 × 3. When putting the model in action, spectrograms of streamed EMG signals are computed on the fly and a sliding window of length $L = 32$ feeds the latest data to the model for real-time inference. Detailed model architecture, data flow and input/output dimension at each layer are illustrated in Figure 4.

Joint classification and regression. Estimating continuous finger-wise forces, by its nature, is a regression problem, and it is natural to adopt common regression losses, such as L_1 or L_2 loss, as the objective function. However, the muscle-generated forces in interactive scenarios have two unique patterns: humans apply forces only sparsely in the real-world; and the variance of force levels is commonly high, ranging from light touches to hard pushes. In practice, regression-based learning oftentimes tends to predict non-zero values (false positive when we do not generate forces) or over-smooth low-amplitude values (false negative for light forces). For our targeted VR/AR applications, this seemingly small estimation error can lead to visually noticeable artifacts and largely compromise user experience (e.g., causing constant vibrations on objects or producing no reaction on low-force touches).

A naïve solution to this problem is to set a cut-off threshold such that the estimated values below it are treated as zero. Although this modification enables zero-value output, tweaking the threshold can be unworkable in practice and the performance is still barely satisfactory as will be shown in Section 4.2. To address this issue, we introduce a classification loss to better differentiate between EMG sequences with and without forces. Specifically, for each time frame t and each finger i, the model outputs a value $p_i^t \in [0, 1]$ indicating the probability of that finger applying force at that time frame. A cross entropy loss L_c is employed to train the model for this force/no-force binary classification task. On top of p_i^t, we compute the predicted force as $\hat{F}_i^t = 2F_{\max} \cdot \max(0, p_i^t - 0.5)$, where $F_{\max}$ denotes the force upper-bound and defines the predicted force range. A L_2 loss is then employed to train the model for force regression.

$$L_c = \frac{1}{T \cdot I} \sum_{t=1}^{T} \sum_{i=1}^{I} y_i^t \cdot \log p_i^t + (1 - y_i^t) \cdot \log(1 - p_i^t) \tag{1}$$

$$L_r = \frac{1}{T \cdot I} \sum_{t=1}^{T} \sum_{i=1}^{I} \|\hat{F}_i^t - F_i^t\|^2 \tag{2}$$

where y_i^t and F_i^t denote the ground-truth force label and value, respectively, for finger i at time frame t.

A hyper-parameter λ is introduced to balance between classification and regression, and the overall loss L takes the form:

$$L = L_c + \lambda \cdot L_r \tag{3}$$

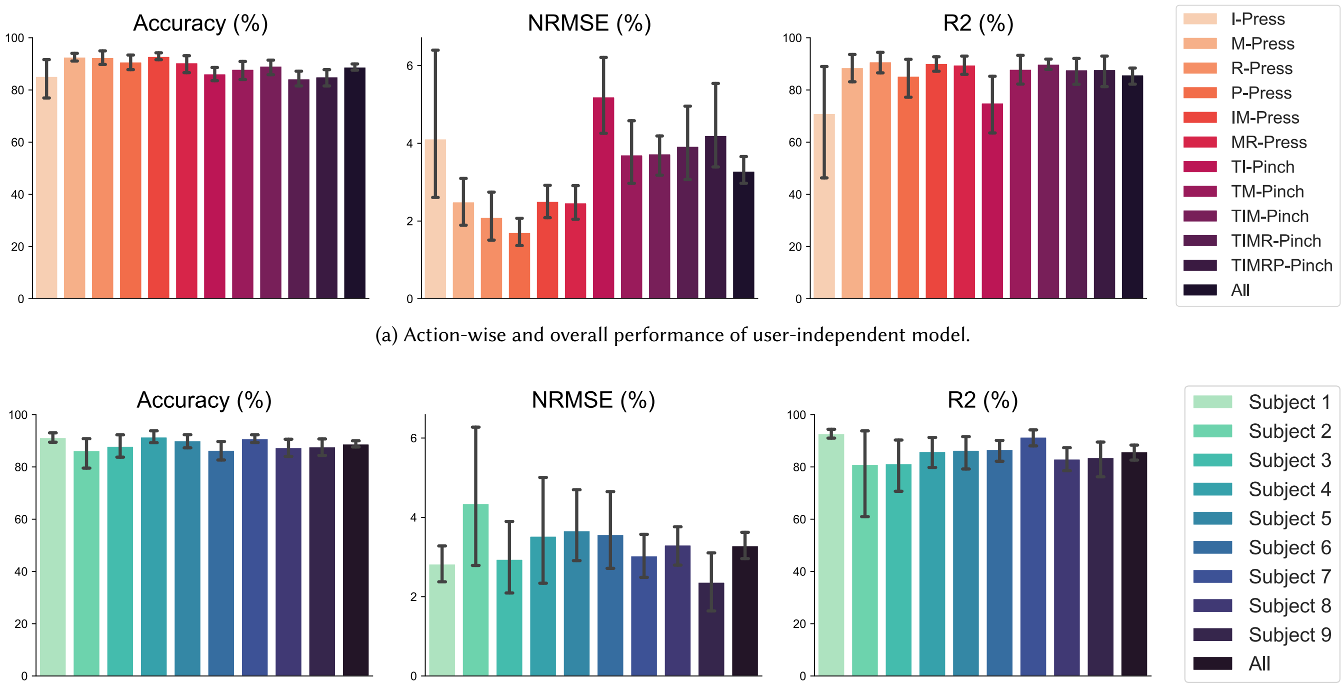
The joint loss above is designed such that, for finger i at time frame t: when $p_i^t \in [0, 0.5]$, we have $\hat{F}_i^t = 0$, only L_c takes effect and the model focuses on correcting wrong classifications; when $p_i^t \in (0.5, 1.0]$, we have $\hat{F}_i^t = 2F_{\max} \cdot (p_i^t - 0.5) \in (0, F_{\max}]$, L_c and L_r together push the model towards the joint classification-regression goal. As a result, we are able to not only get zero-value outputs when there is no force, but also prioritize classification over regression at the beginning stage of training, since the estimated force value will be useless if the predicted class is wrong in the first place. Note that p_i^t is exploited to both differentiate between no-force and force (classification with threshold $p_i^t = 0.5$) and compute the predicted forces $\hat{F}_i^t$ (regression).

4 EVALUATION

To evaluate our method and system, we first discuss in Section 4.1 the specifics of our EMG-Force dataset and the evaluation metrics for quantifying the performance of our CNN-based regression model (detailed in Section 3.3). Then, we present the results of fingertip force estimation for various common hand-object interactions in Section 4.2. Following that, we compare our approach with existing vision-based methods in Section 4.3. We further study the time-efficient generalization of the pre-trained model to new users in Section 4.4. In addition, we also analyze the resulting neural interface in terms of latency and storage for real-time applications in Section 4.5. Finally, we conduct a user study to demonstrate the knowledge of contact force value could benefit material perception and enhances physical realism for real-world VR/AR interaction in Section 4.6.

4.1 EMG-Force Dataset and Evaluation Metrics

The relationship between forearm muscle activations and finger actions exhibits a highly complex mapping [Farina and Holobar 2016]. On top of this complexity, its variations across subject identity, arm/hand posture, and subject's physical condition further add to the complexity of its precise characterization. Additionally, the



(a) Action-wise and overall performance of user-independent model.

(b) Subject-wise and overall performance of user-independent model.

Fig. 6. *Objective evaluation of user-independent model.* (a) shows the action-wise performance of user-independent model in estimating finger-wise forces, with 95% confidence intervals overlaid. Similarly, (b) shows the subject-wise performance.

electric signal detected by each EMG sensor is inevitably a superposition of multiple muscles' activities (as shown in Figure 2), which only makes decoding finger-wise forces from EMG signals even more difficult. Therefore, it is crucial to establish a comprehensive training dataset covering common and natural hand-object contact patterns, so that the neural network model can effectively capture this relationship and acquire better generalization capability. Most prior art relies on the NinaPro dataset [Atzori and Müller 2015] which is actually intended for manipulating robotic arms and is collated from the CyberGlove data glove (and therefore not representative of natural hand or finger movements). Here, we introduce an alternative data set that we have collected ourselves.

EMG-Force dataset. When users perform hand-object interactions, whether in the physical or virtual world, pressing, pushing, pinching, and holding are arguably among the most frequent actions [Ingram et al. 2008]. These actions allow users to not only better perceive surrounding objects, especially their physical properties, but also pick them up for further interactions. Based on their respective force exertion mechanism, we partitioned these actions into two representative groups: pressing/pushing and pinching/holding. Eleven common finger combinations were selected for data collection purposes, with six for the former and five for the latter. This set of actions, which we call action set $\mathcal{A}$, is summarized in Figure 5. To build our EMG-Force dataset, we recruited 9 participants (ages 21 − 30, 5 females, 4 males). Following the data collection and preprocessing pipeline described in Section 3.2, we conducted three collection sessions with each subject, capturing 30 seconds of data for each action during each session. In total, each participant contributed 990-seconds of time-synchronized EMG and force data. For each subject, 2 randomly selected sessions (out of 3) were earmarked for the construction of the training set. The remaining session was withheld and only used for evaluation. When performing pinching actions during data collection, participants were instructed to keep their four fingers over the trackpad and their thumbs below the table, so that they could pinch the ensemble of trackpad and table in a natural manner. Besides, they kept the resultant force stable and balanced (i.e., ∼ 0N). The ground-truth forces for the four fingers were directly recorded by the trackpad, and the force for the thumb was derived as the additive inverse. With the importance of data coverage in mind, all subjects were instructed to randomize their force intensity level within the natural range of each action. In addition, a random spacing in time was enforced between adjacent interactions, so that neural network models do not over-fit to unintended temporal features. The EMG-Force dataset contains light touch less than 1N and firm press up to 30N, covering the typical functional force range of human fingers [Xu et al. 2020b]. In particular, the maximum force for the five fingers in Newton, from the thumb to the pinky finger, are 29.8, 24.4, 25.6, 20.4, and 14.7. The mean/standard deviation/interquartile range are 10.1/8.1/13.9, 5.8/4.6/7.4, 5.7/4.7/7.6,

Table 1. Performance comparison with regular regression losses L_1 and L_2 as well as the no-smoothing variant of our method.

Metric	L_1/L_2 regression	No Smoothing	Ours
Accuracy	85.68% / 85.12%	88.83%	**88.83%**
NRMSE	4.56% / 4.34%	4.02%	**3.29%**
R^2	81.89% / 82.21%	83.59%	**85.82%**

3.4/3.1/4.0, and 2.6/2.5/3.2. Our CNN model predicts force values in $[0, 30]$, which is configured through the force upperbound $F_{\max}$.

Evaluation metrics. To assess the performance of our CNN model in estimating fingertip forces, we adopted three quantitative metrics: (1) Classification Accuracy; (2) Normalized Root Mean Squared Error (NRMSE); (3) Coefficient of Determination, R^2. The model's performance in determining whether a finger exerts force or not at a particular time frame is evaluated by the classification metric, and we only count the model's predictions for a time frame as correct if all five fingers are correctly classified. Using the same notations from Section 3.3, we have:

$$\mathrm{NRMSE} = \frac{1}{F_{\max}}\sqrt{\frac{1}{T \cdot I}\sum_{t=1}^{T}\sum_{i=1}^{I}(F_i^t - \hat{F}_i^t)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{I}(F_i^t - \hat{F}_i^t)^2}{\sum_{t=1}^{T}\sum_{i=1}^{I}(F_i^t - \bar{F}_i^t)^2} \tag{5}$$

where $\bar{F}_i^t$ gives the mean of F_i^t.

4.2 Performance of Decoding Finger-Wise Forces

Experimental setup. Before considering how our scheme applies to specific or new users, we first evaluate the performance of our model in a user-independent setting, where a single model is trained and shared by all users who contributed data. In particular, the entire training set was used to optimize the model against the joint loss defined in Equation (3) for 30 epochs. An Adam optimizer [Kingma and Ba 2015] with constant learning rate of $1e-4$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ was adopted. A weight decay factor of $1e-4$ was enforced to mitigate over-fitting. As a post-processing step, we applied a Gaussian filter of window size 10 to the sequence of predicted force values for temporal smoothing. For ablation purposes, we also trained the model using regular L_1 or L_2 regression loss only, and cut off predicted force values below a small pre-defined threshold. We used PyTorch [Paszke et al. 2019] to implement all our models as well as to perform training and evaluation.

Results. The action-wise and subject-wise performance of the user-independent model is summarized in Figure 6. The overall accuracy, NRMSE, and R^2 are 88.83±6.13%, 3.29±1.76%, and 85.82±14.96%, respectively. On the action side, the model has the highest performance for ring finger pressing, with 92.47±3.89% accuracy, 2.10±0.99% NRMSE, 90.84±6.53% R^2, and the lowest for index finger pressing, with 85.21±11.37% accuracy, 4.12±3.04% NRMSE, 70.99±34.95% R^2. On the subject side, one-way repeated measures ANOVA gives $F_{1,8} = 1.28$, $p = 0.26$ for accuracy, $F_{1,8} = 1.18$, $p = 0.32$ for NRMSE, and $F_{1,8} = 0.82$, $p = 0.59$ for R^2, indicating minor utility discrepancy among subjects. Furthermore, the results of our ablation study are shown in Table 1, validating the effectiveness of the proposed joint classification-regression loss and temporal smoothing.

Table 2. Performance comparison with vision-based methods [Fallahinia and Mascaro 2021a, 2020, 2021b] in terms of NRMSE.

Method	Index	Middle	Ring	Mean
[Fallahinia and Mascaro 2020]	6.1%	5.3%	10.1%	6.2%
[Fallahinia and Mascaro 2021a]	6.1%	5.4%	9.0%	5.9%
[Fallahinia and Mascaro 2021b]	5.7%	4.2%	8.2%	4.9%
Ours	**4.7%**	**3.7%**	**2.4%**	**3.7%**

Discussion. The results above demonstrate the feasibility of accurately decoding finger-wise forces from forearm EMG signals and building robust predictive models that can be shared by multiple users. The statistical significance also suggests that the proposed scheme has the potential of being extended beyond an experimental setting and to more general application scenarios. In addition, it is worth noting that such performance holds under the existence of real-world challenges, such as variations across users in sensor positioning, forearm muscle size, forearm hair thickness, etc. The model is resilient to various discrepancies among users, capable of capturing generalizable EMG-to-force patterns, and achieves utility fairness for users, as evidenced by the ANOVA analysis above, all the while maintaining favorable overall performance.

While these results are statistically rewarding, a remarkably large amount of data is required from each user to support satisfactory performance in practice. Specifically, each participant contributed 22 EMG-Force joint sequences to the training set for the experiment above, which amount to 11 minutes of data. Consider also that there are other inevitable preparations, such as device setup, session break, data pre-processing, and model training. Such delay may become a roadblock for many VR/AR applications in practice. To deploy our neural interface in consumer-level applications, more time-efficient training is essential. This aspect will be addressed in Section 4.4.

4.3 Comparison with Vision-Based Methods

Prior works in the literature have explored vision-based solutions to body force estimation, such as inferring contact forces from the dynamics of hand-object interactions using RGB videos [Ehsani et al. 2020b; Hwang and Lim 2017; Pham et al. 2015, 2017; Zhu et al. 2016] and predicting finger-level forces from the color changes in fingernail imaging [Fallahinia and Mascaro 2021a, 2020, 2021b; Grieve et al. 2010, 2015a,b; Sun et al. 2008]. Similar to our approach, a major benefit of vision-based solution is to bypass on-hand force sensing units. Among these solutions, those based on fingernail imaging also have the potential of delivering accurate and flexible per-finger force estimation for VR/AR applications involving complex hand-object interactions. In this experiment, we compare the accuracy and robustness between our method and three recent vision-based methods [Fallahinia and Mascaro 2021a, 2020, 2021b].

Experimental setup. Due to the challenges of reproducing the identical hardware prototype of data acquisition as in [Fallahinia and Mascaro 2021a, 2020, 2021b], we evaluated our method under the setting adopted by them and compared with their reported

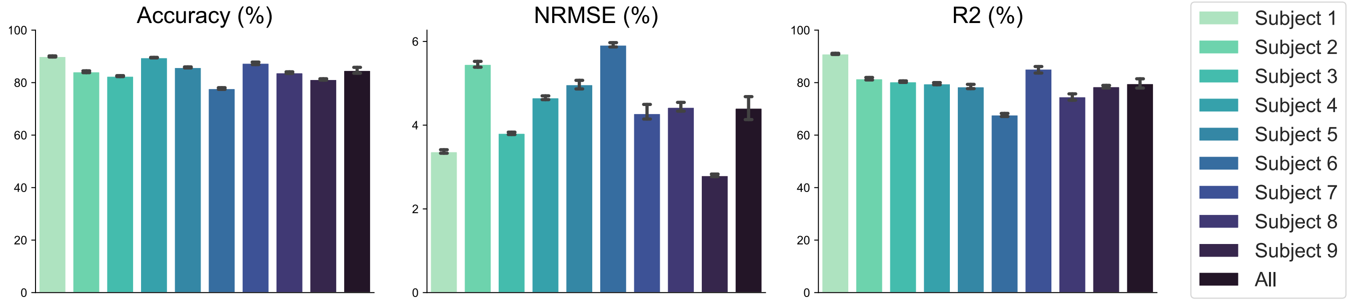


Fig. 7. *Objective evaluation of user-specific model.* Each user-specific model is calibrated using 165-second synchronized EMG-force data.

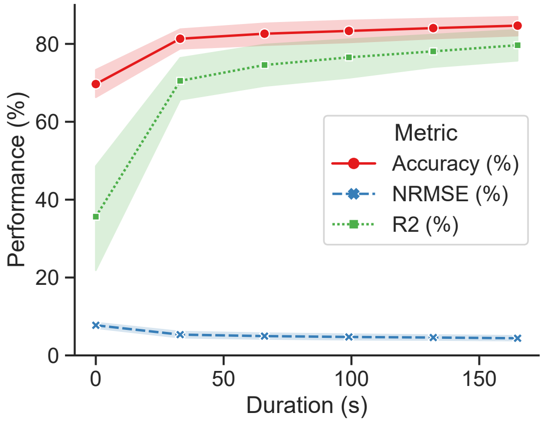


Fig. 8. *Data-efficient calibration of user-specific model via transfer learning.* The performance of user-specific model, as measured by classification accuracy, NRMSE and R^2, is plotted in function of the amount (i.e., time duration) of data collected from the new user for transfer learning. The translucent band around each curve gives the 95% confidence interval.

performance metrics. In particular, [Fallahinia and Mascaro 2021a, 2020, 2021b] only considered single-finger grasping actions and evaluated their method using the index, middle, and ring fingers. To accommodate their evaluation setting, we separated out the partition corresponding to these three single-finger actions from our EMG-Force dataset, i.e., 3x30 seconds (three sessions) of time-synchronized EMG and force data for each subject and each of the three fingers. Two randomly selected sessions were used to train a single user-independent model, while the remaining session was used to evaluate the trained model.

Results. The finger-wise and overall performance of force estimation, as evaluated by NRMSE, is shown in Table 2. Our method outperforms the three vision-based baselines by 2.5%, 2.2%, and 1.2% NRMSE on average, respectively. The advantage is especially noticeable in performance for the ring finger, with our method's NRMSE being less than a quarter of [Fallahinia and Mascaro 2020] and a third of [Fallahinia and Mascaro 2021a,b].

Discussion. Compared to our EMG-based solution, which actively decodes finger-level forces from the controlling muscles' activities, vision-based methods rely on passive observations and are thus highly sensitive to the variations in external factors, such as ambient occlusions, viewing angles, and lighting conditions. Such degrading effects are more problematic for applications involving complex hand-object interactions or real-wild scenarios. On the contrary, our EMG-based solution has shown high robustness to such mentioned issues by its nature. Besides the vulnerability to environmental factors, fingernail-imaging-based methods are also limited in their functional force range, since the variations in fingernail color get less and less detectable as the force intensity increases. The typical functional range for these methods, as evaluated in [Fallahinia and Mascaro 2021a, 2020, 2021b], is around 10N. By contrast, our EMG-based solution is more scalable in terms of force intensity and can robustly estimate forces up to 30N.

4.4 Individualization and Generalization

The analysis of user-independent training in Section 4.2 reveals the need for *time-efficient generalization.* In this section, we investigate how this goal is achievable by extending a pre-trained model for new users using only minimal amounts of data from them. This procedure is commonly referred to as calibration or individualization in human research. Calibration is crucial for machine learning on EMG data since large natural variations exist among different people's muscle-to-EMG patterns. Such discrepancies lead to the well-known generalization challenge that a machine-learned model trained on EMG data commonly fails if directly applied to an unseen user [Phinyomark and Scheme 2018]. Therefore, we decide to perform transfer learning to reduce the data requirements for deploying our model to a new user while maintaining satisfactory prediction performance.

Experimental setup. To evaluate time-efficient generalization via transfer learning, we adopted an experimental setup similar to cross-validation practices for machine learning tasks. Specifically, we first treated subject 1 (S1) as the new user and optimized the model using two sessions of data from each of the other eight subjects. Next, we fine-tuned the resulting model using a portion of data randomly selected from S1's first session, varying from 10% to 50%, to calibrate it into a user-specific model dedicated to S1. Note that 10% session corresponds to 33-second data. Adam optimizer [Kingma and Ba

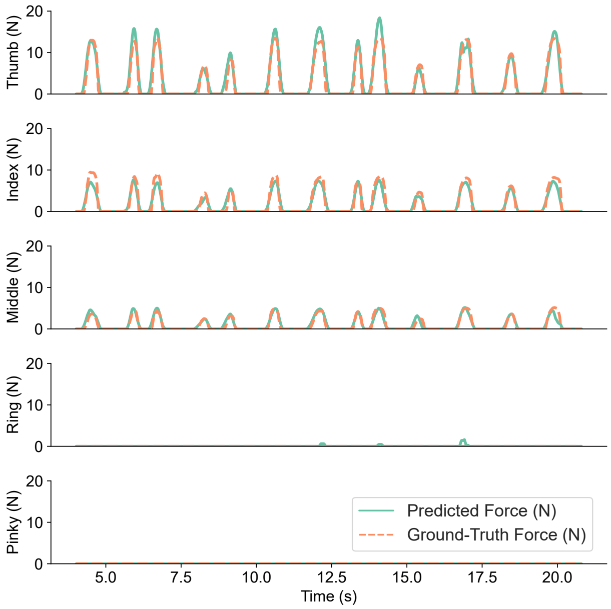


Fig. 9. *Comparison between predicted finger-wise forces and ground truth.* The model is calibrated for subject 8 using only 165-second data from one of his training session and evaluated on a randomly selected sequence from his evaluation session.

2015] with constant learning rate of $5e-5$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$ was adopted. A weight decay factor of $1e-4$ was enforced to mitigate over-fitting. We cycled through the role of new user with each subject to complete the experiment.

Results. The performance of a user-specific model transferred using 165-second data (50% of each subject's first session), as measured by classification accuracy, NRMSE, and R^2, is summarized in Figure 7. At least 81.23% accuracy was consistently observed for all subjects' calibrated models except S6, whose model showed 77.77% accuracy. S1/S4's models achieved over 89.54% accuracy, surpassing the overall performance of the user-independent model with much less training data. The NRMSE metric revealed larger gaps across subjects, which ranged from 2.80% to 5.91%. R^2 is mostly above 75%, with the exception of S6's model yielding 67.73%.

To investigate the minimal amount of data required for calibrating the model towards reasonable utility in practical applications, we analyzed the trade-off between data volume for calibration and resulting model's performance. Figure 8 visualizes the calibrated models' average performance gain as a function of EMG-Force sequences' total length in time. All subjects' models demonstrated rapid improvements as the calibration kicked off and attained 83.33±4.47% accuracy, 4.74±1.09% NRMSE, and 76.55±7.88% R^2 with 66-second data only. The growth rate of mean performance then slowed down, and accuracy/NRMSE gradually plateaued when the duration of data exceeded 150 seconds. Remarkably, these results verify the data efficiency of transfer-learning-based individualization, as evidenced by various accuracy metrics. Taking S8's calibrated model (using 99-second data) as an example, we show a visual comparison between model-predicted and hardware-sensed force values for a randomly selected EMG sequence from their evaluation session in Figure 9. Despite the variations in force intensity and temporal spacing, predicted force values generally aligned well with the ground truth, except for a few slight false positives for the ring finger.

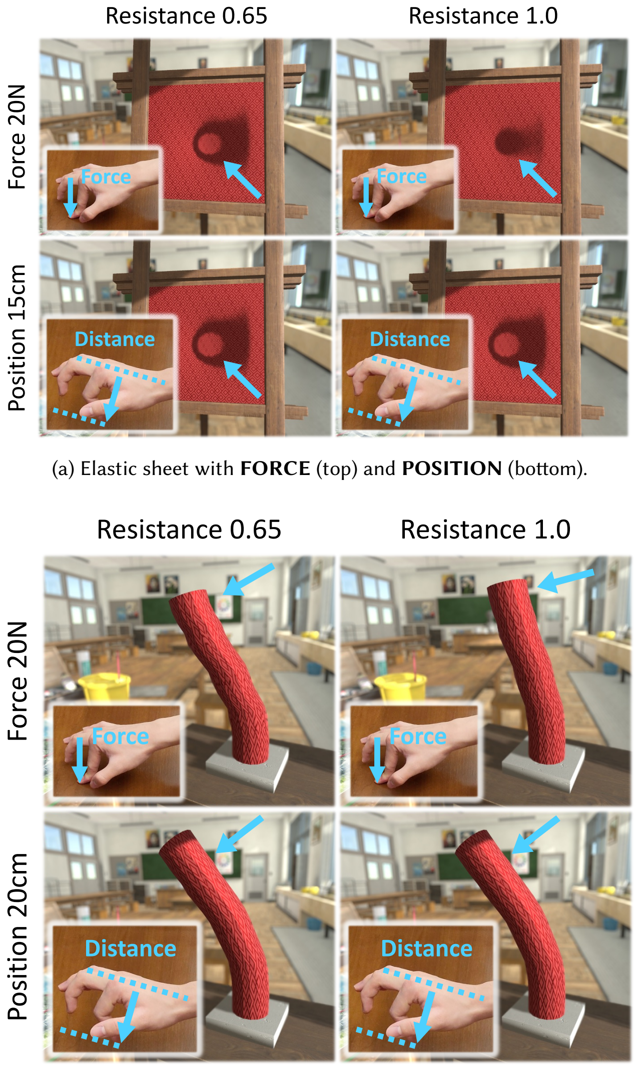
Discussion. Exploiting transfer learning techniques, we demonstrated the feasibility of effectively adapting existing models into dedicated ones for previously unseen users using very limited data from them. Further, as indicated by the above analysis and visualized in Figure 8, the high-precision generalization is achieved with simple and rapid (less than 2 minutes) calibration for novel users.

These findings also circle back to our goal of improving physicality for VR/AR environments in two significant ways. First, our system can harness users' natural abilities and predilections for manipulating things they encounter, thereby expanding the space for developers to create VR/AR experiences that map to real-world scenarios and user behaviors. Importantly, we show that this is achievable for any user, with minimal retooling. Second, as shown in the following section, our system is responsive with low latency. This is crucial, in particular, if we consider that the visual components of AR/VR now routinely refresh at 90 Hz. Any system designed to communicate bodily forces with the virtual environment needs to be agile in timing and nimble in response to data.

4.5 System Performance

Thanks to the moderate scale and bottleneck design of our CNN model, our system's storage and computing requirements are minor for modern PC hardware. The compact model only contains 1.26M 32-bit floating point parameters (≃5MB memory). For the EMG sequence streamed at a particular time frame, i.e., 1248 eight-channel EMG samples, our model only generates around 29.19M Multiply-Accumulate Operations (MACs). Note that while our model requires an EMG sequence of 1248 samples (624ms) as input, these data are only retrieved from history to make predictions for the current time step. When applied in practice, our system's latency performance has two important interfering dimensions: 1) *runtime speed*, i.e., the time needed to complete the force predictions on an EMG sequence of 1248 samples; 2) *reaction latency*, i.e., the duration between EMG data generation and force prediction.

As an implementation detail, we performed GPU parallelization with two consecutive EMG sequences that differ by 32 samples and achieved ≃1.2-1.4ms inference time using a GTX TITAN XP GPU, thus obtaining an approximate 0.7ms latency (i.e. over 1000FPS). Note that this result characterizes the system itself rather than the force output, which is also determined by the spectrogram frame rate. With this setting, the system shall wait for both sequences to arrive, introducing an additional 16ms (32 EMG samples) idle time. That is, although GPU parallelization accelerates runtime speed, it also introduces extra reaction latency. Together with the wireless EMG data transmission latency (2ms) and model inference (0.7ms), our system achieves an overall reaction latency of ≃18.7ms, sufficient for most real-time VR/AR applications [Dangxiao et al. 2019].



(a) Elastic sheet with **FORCE** (top) and **POSITION** (bottom).

(b) Elastic rod with **FORCE** (top) and **POSITION** (bottom).

Fig. 10. *Visualization of the stimuli used in our psychophysical experiment.* (a) illustrates the reaction of two identically-looking elastic sheets differing in material stiffness to inputs from **FORCE** (top) and **POSITION** (bottom). Compared with **POSITION**, the two objects exhibit more natural and prominent difference given the same inputs from **FORCE**. (b) shows similar results for an elastic rod. 3D asset credits to SbbUtutuya and Virtual Method at Unity.

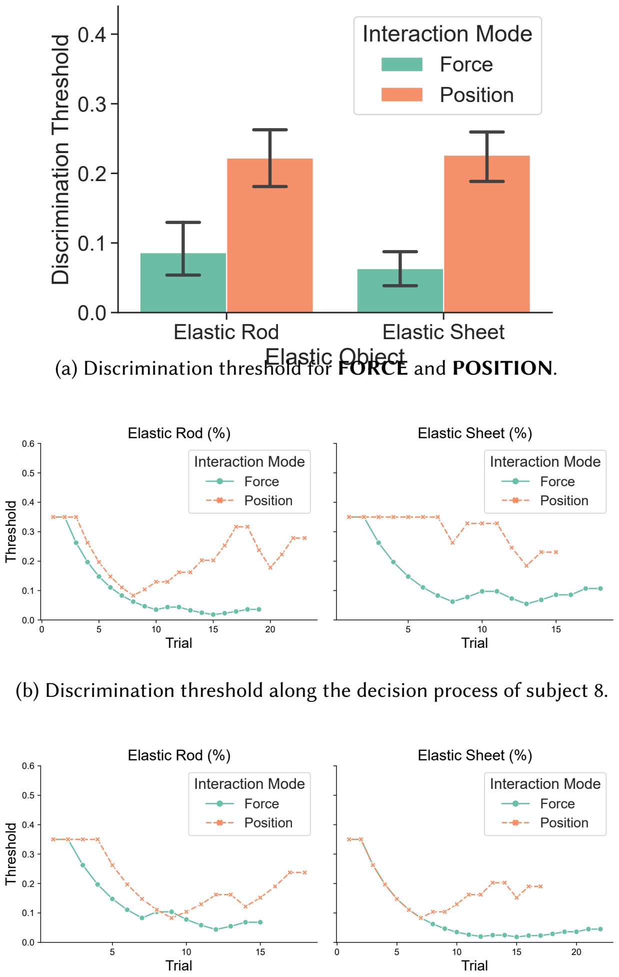
4.6 Psychophysical Study: Enhancing Material Perception in Virtual Environments

A key aim of VR/AR is to create virtual experiences for users as if they were in a physical environment. When interacting with objects in the physical world, we perceive their material properties, such as elasticity and stiffness, through a combination of haptic and visual feedback [Baumgartner et al. 2013]. To recreate such perceptual realism in virtual environments, it is essential to precisely drive virtual objects' motions and deformations via users' muscular forces. We hypothesize that interfaces with such capability may significantly enhance human perception of virtual objects' material properties. In this study, we evaluate to what extent our system, as a real-time force-aware interface, advances toward this goal.

Participants, setup, and calibration. We recruited 12 subjects (ages 20-35, 6 female) to participate in the study. A calibration procedure was performed for each subject before starting the experiment by collecting 1 minute of EMG-Force data from him/her to customize the pre-trained user-independent model (as described in Section 4.4). This calibrated model was then used to estimate finger-wise forces on the EMG signals sensed from that subject in real-time. Estimated force values were communicated to a Unity program via the ZeroMQ library. This pipeline allows for direct application of estimated forces to virtual objects in real-time via physical simulation. During the study, the subjects, wearing an Oculus Quest 2 head-mounted display, remained seated and were free to observe a virtual scene. They interacted with virtual objects in their field of view through unconstrained movements of their forearms, hands, and fingers.

Stimuli. As shown in Figure 10, the visual stimuli were two geometric primitives (elastic sheet and rod) that are soft and deformable. To enable real-time softbody simulation for low-latency interaction on portable VR headsets, we employed an efficient XPBD [Macklin et al. 2016; Müller et al. 2007] implementation by *Virtual Method Studio* [Méndez and Martínez 2021] and only used low-resolution particle models of the virtual objects. It should be noted that simulators' efficiency is orthogonal to the accuracy of model-predicted muscular forces. Therefore, our model can be readily incorporated into *any* simulation system. We adopted deformation resistance of elastic materials as the proxy to represent stiffness with the range $[0, 1]$. Deformation resistance measures a physical material's ability to resist externally loaded forces. In this study, we aim to identify participants' discriminative thresholds of virtual objects' stiffness under varying conditions. To avoid visual cues biasing the results, all objects were rendered with identical material and texture, regardless of their physical properties.

Conditions. For evaluation purposes, all subjects were instructed to employ two interaction methods in sequence during the study (the order was random). Besides our data-driven system for force-enabled interaction (**FORCE**), we also included position-based interaction (**POSITION**) for comparison. Specifically, **POSITION** is a commonly adopted solution in commercial VR/AR systems that lacks force information. It allows users to modify virtual objects' position, orientation, and shape by colliding their hands with the objects. For **FORCE**, users interacted with virtual objects by manipulating physical proxies while our system estimated their muscular forces. These forces were then used to deform the two virtual objects, including indenting the sheet's center and bending the rod's top. For **POSITION**, we leverage the hand tracking capability of Oculus Quest 2 to estimate the flexion level of users' index finger (distance from the index fingertip to the palm plane) and use it as input for the identical interaction as **FORCE**. Specifically, this value determines the indentation depth for the sheet's center and the bending level for the rod's top. Figure 10 illustrates these interactions. Notably,



(a) Discrimination threshold for **FORCE** and **POSITION**.

(b) Discrimination threshold along the decision process of subject 8.

(c) Discrimination threshold along the decision process of subject 10.

Fig. 11. *Psychophysical study on material perception in VR.* (a) shows 12 subjects' discrimination thresholds on two geometric primitives for **FORCE** and **POSITION**. The error bars indicate 95% confidence intervals. A remarkably and consistently lower threshold for **FORCE** can be observed. (b) and (c) visualize the discrimination threshold along the decision process of two subjects when they were engaging in our 1-up-2-down staircase protocol with 2AFC trials (which object is stiffer). Note the consistent decreasing trend of **FORCE**, indicating high confidence level when the subject made 2AFC decisions.

while physical proxies that resemble the virtual objects will enhance users' experience with **FORCE**, **POSITION** does not benefit from them. To avoid users favoring **FORCE** due to irrelevant features of the physical proxies, we intentionally used a hard and flat table.

Task. To measure participants' (perceptual) discriminative threshold of material stiffness, we employed a psychophysical task as a 2-alternative-forced-choice (2AFC) with a 1-up-2-down staircase procedure (5-reversal to confirm convergence thus termination). For softer and stiffer objects, the experiment started with deformation resistance equal to 0.65 and 1.0, i.e. a threshold of 0.35. Each time the threshold got updated, it was incremented or decremented by a quarter of the current threshold, and the two deformation resistance values were updated such that their mean was unchanged. Specifically, the two interaction conditions (**FORCE**/**POSITION**) were sequentially presented to the user for consideration (with a random and counter-balanced order). During the experiment, the participants were instructed to freely interact with the corresponding stimuli and then indicate (using a keyboard) which one of the two stimuli appeared stiffer. They observed the deformation pattern along with the proactive intervention. After each trial, the participants chose one of the two stimuli that appeared stiffer. Each 2AFC trial took 5 seconds. A warm-up session was first performed to allow each individual user to familiarize with the stimuli and the interaction design. For each participant, the entire experiment took about half an hour. The number of trials (ranging from 25 to 38) depended on the speed of the staircase convergence.

Metrics and results. Figure 11 visualizes our statistical results. The mean discriminative thresholds of **FORCE**/**POSITION** were 0.09±0.07, 0.22±0.07 for the elastic rod, and 0.06±0.04, 0.23±0.07 for the elastic sheet, indicating 61.3% and 72.1% improvements with **FORCE**, respectively. One-way repeated measures ANOVA shows that the effects of interaction method are statistically significant: $F_{1,11} = 20.99$ $p = 1.46e^{-4}$ for the elastic rod and $F_{1,11} = 46.79$ $p = 7.16e^{-7}$ for the elastic sheet.

Discussion. We designed our psychophysical experiment to test whether users could quantitatively perceive realistic soft objects and their material properties in AR/VR. We used primitive geometries and their natural articulated abilities to explore, examine, and assess them through force-based interaction with the forearm, hand, and fingers. In other words, we wished to test whether users could simply enter a VR/AR scene, start to prod and poke things in that scene, and leave with a sense that the objects responded with realistic physics.

The results showed a statistically significantly lower discrimination threshold while participants interacted with virtual objects with **FORCE**. That is, the force-visual correlated interaction facilitated significantly more realistic perception of virtual objects' physical characteristics when users were engaged in free manipulation within the virtual world. We regard this as a significant proof of concept for our approach. Consider that, in the real world, humans spend much of their infancy working out how to muster the forces available to them in their arms, hands, and fingers, through ongoing trial and error with the things that they encounter. In essence, we capture the small electrical signals that human muscles cast as they put their skills to use, and we are able to use these signals as indices for machine-learning what that might mean in physics. Asking and answering how our human users believe that the physical response our system yields are realistic-seeming establishes the perceptual foundation of various new possibilities of interfaces.

5 APPLICATIONS

Beyond enhancing the physical realism of hand-object interactions in VR/AR through more natural and intuitive haptic inputs, our

method of decoding forces fron
following *general* application sce
capabilities—representing virtu
virtual control—because they fo
cific VR/AR applications, e.g., g
technology, training, social com

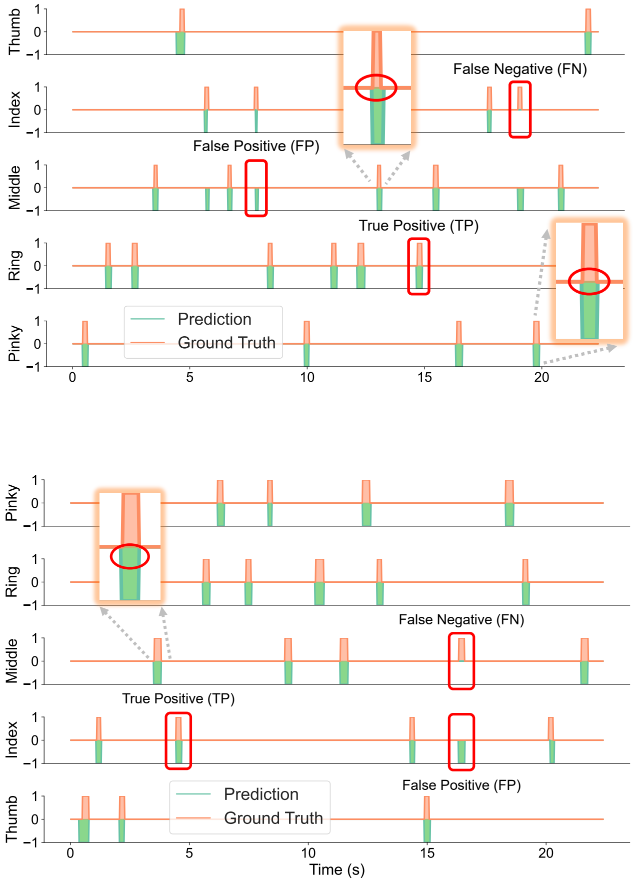
Multimodal data sources for vi
realistic human behaviors in virt
and essential topic [Hassan et al.
action data from optical motion
knowledge of users' forces on tl
is essential in interactive and c
2009] remains missing. Our syste
foundation that collects the mo
actions.

Accessible interfaces. People
face challenges when interacti
interfaces, through actions su
prosthetic controls with neura
promising assistive technolog
posed research may predict the
forearm remotely, allowing fo
interfaces for people with hand
as during driving or cold outdo

Ubiquitous control. The pro
hand interactions and introdu
incurs unnoticeable change to
sides, it is usable in most daily
concerns. As a result, we can l
body input devices for ubiquitous control. For instance, we can map patterns of finger pressing to the buttons in a software's control panel or the keys on a musical instrument. As illustrative of broader applications, in the following, we demonstrate that our method can be readily adapted to perform robust finger identification during multi-finger tapping for ubiquitous control.

5.1 Case Study: Ubiquitous Control via Finger Tapping

Traditional computers typically equip with dedicated control devices such as mouse and keyboard. However, the ultimate goal of VR and AR platforms is a transformative natural and ubiquitous control. To this end, researchers have recently attempted to infer user intention from natural modalities, such as vision [Han et al. 2020; Kim et al. 2012; Stearns et al. 2018], acoustics [Harrison and Hudson 2008; Xu et al. 2020a; Zhang et al. 2018], radar [Lien et al. 2016; Wang et al. 2016], and Wi-Fi [Abdelnasser et al. 2018, 2015]. Despite their support for eyes-free control, these methods are susceptible to environmental interference, such as occlusions and noise, and may suffer from performance decay in complex environments. By contrast, EMG-based solutions electronically tracks users' forearms as input devices. In this experiment, we validate our method's performance while being applied to detect "click" actions, which are identified as any finger's tapping.



(b) Multi-finger tapping (left hand).

Fig. 12. *Finger identification during multi-finger tapping.* (a) and (b) visualize the ground-truth (orange bars) and the identification results by our method (green bars) of two randomly-sampled sequences of multi-finger tapping from the hold-out data. Bar width denotes tapping duration.

Table 3. Tapping finger identification performance.

Metric	Thumb	Index	Middle	Ring	Pinky	Mean
Precision	93.3%	88.8%	84.9%	98.4%	94.4%	92.0%
Recall	94.2%	78.4%	94.7%	100.0%	96.2%	92.7%

Experimental setup. All 10 fingers from both hands are included. One male subject participated in the study. Following the data collection and pre-processing pipeline (Section 3.2), we conducted four collection sessions. During each of the first three sessions, the subject performed single-finger tapping actions and captured 30 seconds of tapping data for each of the 10 fingers. During the last session, the subject performed random multi-finger tapping actions and captured 150 seconds of tapping data for each hand. The subject was instructed to tap a trackpad in a natural unconstrained manner in all four sessions. In total, the subject contributed 1200-second of time-synchronized EMG and force data. The force data was subsequently converted into $\{0, 1\}$ labels, i.e., "tap" and "no-tap". Two randomly selected sessions (out of the first three sessions) were used to construct the training set. The remaining two sessions were withheld and only used for evaluation. The CNN model was trained using Adam optimizer and cross-entropy loss for the task of per-frame finger-wise "tap" or "no-tap" classification for 20 epochs. The learning rate started at $1e-3$ and dropped to $1e-4$ at epoch 10. A weight decay factor of $1e-4$ was enforced to mitigate over-fitting. As a post-processing step, we applied a mean filter of window size 10 to the sequence of predicted tapping probabilities for temporal smoothing. The predicted tapping probability for each time frame was compared with a threshold of 0.3 to determine if it is "tap".

Metrics. We adapted two evaluation metrics from machine learning, precision (P) and recall (R), to accommodate our experimental setting. Here, P/R denotes the proportion of correctly detected tapping among all detected/ground-truth tapping. Note that both metrics are evaluated in a finger-wise manner.

$$\mathrm{P}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i}; \quad \mathrm{R}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i}. \tag{6}$$

Here, TP (true positive), FP (false positive), and FN (false negative) denote the number of correctly detected tapping, falsely detected tapping, and undetected ground-truth tapping, respectively. Fingers are indexed by $i \in \{1, 2, 3, 4, 5\}$. In particular, a sequence of consecutive "tap" predictions of duration longer than 0.1 second is considered as a detected tapping. We consider correct detection if the temporal Intersection over Union (IoU) between its interval and the ground-truth is greater than 0.5, and falsely detected otherwise.

Results. As shown in Table 3, our CNN model achieves mean precision of 92.0% and mean recall of 92.7% using only 10-minute data for training. The detection quality for the ring finger is the best, with 98.4% precision and 100.0% recall. The index finger showed lower recall 78.4%, while the middle finger showed lower precision 84.9%. Figure 12 visualizes the ground-truth (orange) and model-detected tappings (green) for two randomly-sampled sequences of multi-finger tapping from the last session, one for each hand. As we can see, most tapping actions are correctly detected by the model, with few false positives and false negatives. Accurate tapping duration and detection latency can also be observed.

Discussion. The above results demonstrate our method's applicability to effectively detect finger tapping as an ubiquitous interface. The definitions of precision and recall inherently establish a trade-off between the two metrics: the higher a model achieves in terms of recall, the lower it gets for precision, and vice-versa. Such trade-off can be translated into the context of this specific task as: the more sensitive the model is in detecting finger tapping (more positive predictions), the more actual finger tapping made by the user it will detect (higher recall), and the more mistakes it will make (lower precision). Conveniently, we can increase or decrease the detection threshold for "tap" to prioritize over precision or recall, depending on the demand in the actual application scenario. For instance, if we are considering an application where the accuracy of control signals outweighs the response rate, we might consider lowering the model's sensitivity to sacrifice a bit of recall for better precision.

6 LIMITATIONS AND FUTURE WORK

In this paper, by leveraging EMG sensors on the forearm, we demonstrate the possibility of tracking, predicting, and transferring human muscular forces in the physical world to interactions in virtual representations and environments. In other words, we open-up pathways for the virtual world to react precisely to human-induced physical forces from the real world. Our objective measurements and subject psychophysical experiments support the framework's robustness, accuracy, generalizability, and real-world benefits in enhancing human perceptual understanding of physical materials. This is achieved by our tailored dataset and real-time deep learning approach on muscular signals.

However, several limitations remain for future investigation. First, the dataset that trains our system (Figure 5) were generated by user actions with an off-the-shelf pad-like force sensor. Consequently, the method does not robustly encode complex hand or full-body poses with higher dimensionality and degree-of-freedom, as would be the case if a user was exerting force on more complicated three-dimensional objects. This could be resolved by incorporating data using recent advancements in wearable sensing devices [Luo et al. 2021a; Sundaram et al. 2019], which could enable broader data coverage and thus free-form interaction with complex geometric shapes. Second, our method is practically generalizable but still requires a short (1 minute) calibration process to ensure high-quality predictions. An exciting future direction is extending the framework with unsupervised learning. We believe an automated individualization mechanism may unlock the potential of a fully adaptive framework for arbitrary users without access to the calibration setup. Third, the muscle signals may show different patterns with active (e.g., clenching fist) and resisting (directly interacting with physical objects) forces. Integrating hand tracking and arm pose data into the model may shed light on differentiating the two means. Lastly, to enable accessibility applications, we plan to extend the data and evaluate the system's performance on a larger population, including people with limb impairments. Indeed the ability to sense hand and finger forces and actions directly from forearm muscle signals could establish VR/AR as an entirely new modality for democratizing access to computer graphics applications across a much broader range of interaction abilities. This, we consider, is where development of force-aware VR/AR could be fantastically useful.

ACKNOWLEDGMENTS

This research is partially supported by the National Science Foundation (NSF) under Grant Nos. 2232817, 2225861, 2027652, and 1729815.

REFERENCES

Heba Abdelnasser, Khaled Harras, and Moustafa Youssef. 2018. A ubiquitous WiFi-based fine-grained gesture recognition system. *IEEE Transactions on Mobile Computing* 18, 11 (2018), 2474–2487.

Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. 2015. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 1472–1480.

Md Rezwanul Ahsan, Muhammad I Ibrahimy, Othman O Khalifa, et al. 2009. EMG signal classification for human computer interaction: a review. *European Journal of Scientific Research* 33, 3 (2009), 480–501.

Deni Andrean, Daniel S Pamungkas, and Sumantri Kurniawan Risandriya. 2019. Controlling Robot Hand Using FFT as Input to the NN Algorithm. In *Journal of Physics:*

Conference Series, Vol. 1230. IOP Publishing, 012030.

Gopala K Anumanchipalli, Josh Chartier, and Edward F Chang. 2019. Speech synthesis from neural decoding of spoken sentences. *Nature* 568, 7753 (2019), 493–498.

Manfredo Atzori and Henning Müller. 2015. The Ninapro database: a resource for sEMG naturally controlled robotic hand prosthetics. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 7151–7154.

Tara Baldacchino, William R Jacobs, Sean R Anderson, Keith Worden, and Jennifer Rowson. 2018. Simultaneous force regression and movement classification of fingers via surface EMG within a unified Bayesian framework. *Frontiers in bioengineering and biotechnology* 6 (2018), 13.

Berj Bardizbanian, Jennifer Keating, Xinming Huang, and Edward A Clancy. 2020a. 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)*. IEEE, 3134–3137.

Berj Bardizbanian, Ziling Zhu, Jianan Li, Xinming Huang, Chenyun Dai, Carlos Martinez-Luna, Benjamin E McDonald, Todd R Farrell, and Edward A Clancy. 2020b. Efficiently training two-DoF hand-wrist EMG-force models. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 369–373.

Elisabeth Baumgartner, Christiane B. Wiebel, and Karl R. Gegenfurtner. 2013. Visual and haptic representations of material properties. *Multisensory Research* 26, 5 (2013), 429–455. https://doi.org/10.1163/22134808-00002429

Vincent Becker, Pietro Oldrati, Liliana Barrios, and Gábor Sörös. 2018. Touchsense: classifying finger touches and measuring their force with an electromyography armband. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. 1–8.

Hrvoje Benko, Christian Holz, Mike Sinclair, and Eyal Ofek. 2016. Normaltouch and texturetouch: High-fidelity 3d haptic shape rendering on handheld virtual reality controllers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 717–728.

Joanna Bergström and Kasper Hornbæk. 2019. Human–Computer Interaction on the Skin. *ACM Computing Surveys (CSUR)* 52, 4 (2019), 1–14.

Reza Boostani and Mohammad Hassan Moradi. 2003. Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiological measurement* 24, 2 (2003), 309.

Adnane Boukhayma, Rodrigo de Bem, and Philip HS Torr. 2019. 3d hand shape and pose from images in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10843–10852.

Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7291–7299.

Claudio Castellini and Risto Koiva. 2012. Using surface electromyography to predict single finger forces. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 1266–1272.

Claudio Castellini and Patrick Van Der Smagt. 2009. Surface EMG in advanced hand prosthetics. *Biological cybernetics* 100, 1 (2009), 35–47.

Lele Chen, Chen Cao, Fernando De la Torre, Jason Saragih, Chenliang Xu, and Yaser Sheikh. 2021. High-fidelity Face Tracking for AR/VR via Deep Lighting Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13059–13069.

Minsu Cho, Younggeol Cho, and Kyung-Soo Kim. 2022. Training Strategy and sEMG Sensor Positioning for Finger Force Estimation at Various Elbow Angles. *International Journal of Control, Automation and Systems* 20, 5 (2022), 1621–1631.

Inrak Choi, Heather Culbertson, Mark R Miller, Alex Olwal, and Sean Follmer. 2017. Grabity: A wearable haptic interface for simulating weight and grasping in virtual reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 119–130.

Inrak Choi, Elliot W Hawkes, David L Christensen, Christopher J Ploch, and Sean Follmer. 2016. Wolverine: A wearable haptic interface for grasping in virtual reality. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 986–993.

Inrak Choi, Eyal Ofek, Hrvoje Benko, Mike Sinclair, and Christian Holz. 2018. Claw: A multifunctional handheld haptic controller for grasping, touching, and triggering in virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

Hang Chu, Shugao Ma, Fernando De la Torre, Sanja Fidler, and Yaser Sheikh. 2020. Expressive telepresence via modular codec avatars. In *European Conference on Computer Vision*. Springer, 330–345.

Wang Dangxiao, Guo Yuan, Liu Shiyi, Yuru Zhang, Xu Weiliang, and Xiao Jing. 2019. Haptic display for virtual reality: progress and challenges. *Virtual Reality & Intelligent Hardware* 1, 2 (2019), 136–162.

Yu Du, Yongkang Wong, Wenguang Jin, Wentao Wei, Yu Hu, Mohan S Kankanhalli, and Weidong Geng. 2017. Semi-Supervised Learning for Surface EMG-based Gesture Recognition.. In *IJCAI*. 1624–1630.

Kiana Ehsani, Shubham Tulsiani, Saurabh Gupta, Ali Farhadi, and Abhinav Gupta. 2020a. Use the Force, Luke! Learning to Predict Physical Forces by Simulating Effects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kiana Ehsani, Shubham Tulsiani, Saurabh Gupta, Ali Farhadi, and Abhinav Gupta. 2020b. Use the force, luke! learning to predict physical forces by simulating effects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 224–233.

Marc O Ernst and Martin S Banks. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 6870 (2002), 429–433.

Navid Fallahinia and Stephen Mascaro. 2021a. Real-Time Tactile Grasp Force Sensing Using Fingernail Imaging via Deep Neural Networks. *arXiv preprint arXiv:2109.15231* (2021).

Navid Fallahinia and Stephen A Mascaro. 2020. Comparison of Constrained and Unconstrained Human Grasp Forces Using Fingernail Imaging and Visual Servoing. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2668–2674.

Navid Fallahinia and Stephen A Mascaro. 2021b. Feasibility study of force measurement for multi-digit unconstrained grasping via fingernail imaging and visual servoing. *ASME Letters in Dynamic Systems and Control* 1, 2 (2021).

Yinfeng Fang, Dalin Zhou, Kairu Li, Zhaojie Ju, and Honghai Liu. 2019. Attribute-driven granular model for EMG-based pinch and fingertip force grand recognition. *IEEE transactions on cybernetics* 51, 2 (2019), 789–800.

Dario Farina and Aleš Holobar. 2016. Characterization of human motor units from surface EMG decomposition. *Proc. IEEE* 104, 2 (2016), 353–373.

Sharlene N Flesher, John E Downey, Jeffrey M Weiss, Christopher L Hughes, Angelica J Herrera, Elizabeth C Tyler-Kabara, Michael L Boninger, Jennifer L Collinger, and Robert A Gaunt. 2021. A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science* 372, 6544 (2021), 831–836.

Alycia Gailey, Panagiotis Artemiadis, and Marco Santello. 2017. Proof of concept of an online EMG-based decoding of hand postures and individual digit forces for prosthetic hand control. *Frontiers in neurology* 8 (2017), 7.

Eric J Gonzalez, Eyal Ofek, Mar Gonzalez-Franco, and Mike Sinclair. 2021. 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*. 732–742.

Thomas Grieve, Lucas Lincoln, Yu Sun, John M Hollerbach, and Stephen A Mascaro. 2010. 3d force prediction using fingernail imaging with automated calibration. In *2010 IEEE Haptics Symposium*. IEEE, 113–120.

Thomas R Grieve, John M Hollerbach, and Stephen A Mascaro. 2015a. 3-d fingertip touch force prediction using fingernail imaging with automated calibration. *IEEE Transactions on Robotics* 31, 5 (2015), 1116–1129.

Thomas R Grieve, John M Hollerbach, and Stephen A Mascaro. 2015b. Optimizing fingernail imaging calibration for 3d force magnitude prediction. *IEEE transactions on haptics* 9, 1 (2015), 69–79.

Paras Gulati, Qin Hu, and S Farokh Atashzar. 2021. Toward Deep Generalization of Peripheral EMG-Based Human-Robot Interfacing: A Hybrid Explainable Solution for NeuroRobotic Systems. *IEEE Robotics and Automation Letters* 6, 2 (2021), 2650–2657.

Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, et al. 2020. MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 87–1.

Chris Harrison and Scott E Hudson. 2008. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 205–208.

Mohamed Hassan, Duygu Ceylan, Ruben Villegas, Jun Saito, Jimei Yang, Yi Zhou, and Michael Black. 2021. Stochastic Scene-Aware Motion Prediction. In *Proceedings of the International Conference on Computer Vision 2021*.

Mamoru Hirota, Ayumu Tsuboi, Masayuki Yokoyama, and Masao Yanagisawa. 2018. Gesture recognition of air-tapping and its application to character input in VR space. In *SIGGRAPH Asia 2018 Posters*. 1–2.

Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick Van Der Smagt, et al. 2012. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485, 7398 (2012), 372–375.

At L Hof. 1991. Errors in frequency parameters of EMG power spectra. *IEEE transactions on biomedical engineering* 38, 11 (1991), 1077–1088.

Ruochen Hu, Xiang Chen, Haotian Zhang, Xu Zhang, and Xun Chen. 2022. A Novel Myoelectric Control Scheme Supporting Synchronous Gesture Recognition and Muscle Force Estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2022).

Wonjun Hwang and Soo-Chul Lim. 2017. Inferring interaction force from visual information without using physical force sensors. *Sensors* 17, 11 (2017), 2455.

James N Ingram, Konrad P Körding, Ian S Howard, and Daniel M Wolpert. 2008. The statistics of natural hand movements. *Experimental brain research* 188, 2 (2008), 223–236.

Sumit Jain, Yuting Ye, and C Karen Liu. 2009. Optimization-based interactive motion synthesis. *ACM Transactions on Graphics (TOG)* 28, 1 (2009), 1–12.

Haider Ali Javaid, Mohsin Islam Tiwana, Ahmed Alsanad, Javaid Iqbal, Muhammad Tanveer Riaz, Saeed Ahmad, and Faisal Abdulaziz Almisned. 2021. Classification of Hand Movements Using MYO Armband on an Embedded Platform. *Electronics* 10, 11 (2021), 1322.

Yong-Un Jo and Do-Chang Oh. 2020. REAL-TIME HAND GESTURE CLASSIFICATION USING CRNN WITH SCALE AVERAGE WAVELET TRANSFORM. *Journal of Mechanics in Medicine and Biology* 20, 10 (2020), 2040028.

Hanbyul Joo, Tomas Simon, and Yaser Sheikh. 2018. Total capture: A 3d deformation model for tracking faces, hands, and bodies. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 8320–8329.

Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. 2018. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 7122–7131.

David Kim, Otmar Hilliges, Shahram Izadi, Alex D Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology.* 167–176.

Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akşit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, et al. 2019. Foveated AR: dynamically-foveated augmented reality display. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15.

Diederik P. Kingma and Jimmy Ba. 2015. 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.*

Babis Koniaris, Ivan Huerta, Maggie Kosek, Karen Darragh, Charles Malleson, Joanna Jamrozy, Nick Swafford, Jose Guitian, Bochang Moon, Ali Israr, et al. 2016. Iridium: immersive rendered interactive deep media. In *ACM SIGGRAPH 2016 VR Village.* 1–2.

Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F Welch, and Gerd Bruder. 2018. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–11.

Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. 2016. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–19.

Ming C Lin and Miguel Otaduy. 2008. *Haptic rendering: foundations, algorithms, and applications.* CRC Press.

Pu Liu, Donald R Brown, Edward A Clancy, Francois Martel, and Denis Rancourt. 2013. EMG-force estimation for multiple fingers. In *2013 IEEE Signal Processing in Medicine and Biology Symposium (SPMB).* IEEE, 1–6.

Yilin Liu, Shijia Zhang, and Mahanth Gowda. 2021. NeuroPose: 3D Hand Pose Tracking using EMG Wearables. In *Proceedings of the Web Conference 2021.* 1471–1482.

Yiyue Luo, Yunzhu Li, Pratyusha Sharma, Wan Shou, Kui Wu, Michael Foshey, Beichen Li, Tomás Palacios, Antonio Torralba, and Wojciech Matusik. 2021a. Learning human–environment interactions using conformal tactile textiles. *Nature Electronics* 4, 3 (2021), 193–201.

Yiyue Luo, Kui Wu, Tomás Palacios, and Wojciech Matusik. 2021b. KnitUI: Fabricating interactive and sensing textiles with machine knitting. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–12.

Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. 2021. Pixel Codec Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 64–73.

Miles Macklin, Matthias Müller, and Nuttapong Chentanez. 2016. XPBD: position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games.* 49–54.

He Mao, Peng Fang, and Guanglin Li. 2021. Simultaneous estimation of multi-finger forces by surface electromyography and accelerometry signals. *Biomedical Signal Processing and Control* 70 (2021), 103005.

Itzel Jared Rodríguez Martínez, Andrea Mannini, Francesco Clemente, and Christian Cipriani. 2020. Online grasp force estimation from the transient EMG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 10 (2020), 2333–2341.

Itzel Jared Rodriguez Martinez, Andrea Mannini, Francesco Clemente, Angelo Maria Sabatini, and Christian Cipriani. 2020. Grasp force estimation from the transient EMG using high-density surface recordings. *Journal of Neural Engineering* 17, 1 (2020), 016052.

W Melzer, E Rios, and MF Schneider. 1984. Time course of calcium release and removal in skeletal muscle fibers. *Biophysical journal* 45, 3 (1984), 637–641.

Matthias Müller, Bruno Heidelberger, Marcus Hennix, and John Ratcliff. 2007. Position based dynamics. *Journal of Visual Communication and Image Representation* 18, 2 (2007), 109–118.

Jose María Méndez and Lidia Martínez. 2021. Virtual Method Studio. Retrieved 2021-12-01 from http://obi.virtualmethodstudio.com/index.html

Alejandro Newell, Kaiyu Yang, and Jia Deng. 2016. Stacked hourglass networks for human pose estimation. In *European conference on computer vision.* Springer, 483–499.

Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* (2016).

Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality* 23, 2 (2019), 119–131.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019), 8026–8037.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.

Tu-Hoa Pham, Abderrahmane Kheddar, Ammar Qammaz, and Antonis A Argyros. 2015. Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2810–2819.

Tu-Hoa Pham, Nikolaos Kyriazis, Antonis A Argyros, and Abderrahmane Kheddar. 2017. Hand-object contact force estimation from markerless visual tracking. *IEEE transactions on pattern analysis and machine intelligence* 40, 12 (2017), 2883–2896.

Angkoon Phinyomark and Erik Scheme. 2018. EMG pattern recognition in the era of big data and deep learning. *Big Data and Cognitive Computing* 2, 3 (2018), 21.

S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori. 2017. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLoS One* 12, 10 (2017), e0186132. https://doi.org/10.1371/journal.pone.0186132

Wen Qi, Hang Su, Junhao Zhang, Rong Song, Giancarlo Ferrigno, Elena De Momi, and Andrea Aliverti. 2021. Active Learning Strategy of Finger Flexion Tracking using sEMG for Robot Hand Control. In *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM).* IEEE, 753–758.

Elahe Rahimian, Soheil Zabihi, Amir Asif, S Farokh Atashzar, and Arash Mohammadi. 2021. Few-Shot Learning for Decoding Surface Electromyography for Hand Gesture Recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 1300–1304.

Javier Romero, Dimitrios Tzionas, and Michael J Black. 2017. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics (ToG)* 36, 6 (2017), 1–17.

Farshid Salemi Parizi, Wolf Kienzle, Eric Whitmire, Aakar Gupta, and Hrvoje Benko. 2021. RotoWrist: Continuous Infrared Wrist Angle Tracking using a Wristband. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology.* 1–11.

Christopher Spiewak, M Islam, A Zaman, and Mohammad Habibur Rahman. 2018. A comprehensive study on EMG feature extraction and classifiers. *Open Access Journal of Biomedical Engineering and Biosciences* 1, 1 (2018), 1–10.

Shriya S Srinivasan, Samantha Gutierrez-Arango, Ashley Chia-En Teng, Erica Israel, Hyungeun Song, Zachary Keith Bailey, Matthew J Carty, Lisa E Freed, and Hugh M Herr. 2021. Neural interfacing architecture enables enhanced motor control and residual limb functionality postamputation. *Proceedings of the National Academy of Sciences* 118, 9 (2021).

Lee Stearns, Uran Oh, Leah Findlater, and Jon E Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–23.

Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan McGuire, David Luebke, and Arie Kaufman. 2018. Towards virtual reality infinite walking: dynamic saccadic redirection. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.

Tianyun Sun, Qin Hu, Jacqueline Libby, and S Farokh Atashzar. 2022. Deep Heterogeneous Dilation of LSTM for Transient-phase Gesture Prediction through High-density Electromyography: Towards Application in Neurorobotics. *IEEE Robotics and Automation Letters* (2022).

Yu Sun, John M Hollerbach, and Stephen A Mascaro. 2008. Predicting fingertip forces by imaging coloration changes in the fingernail and surrounding skin. *IEEE Transactions on Biomedical Engineering* 55, 10 (2008), 2363–2371.

Subramanian Sundaram, Petr Kellnhofer, Yunzhu Li, Jun-Yan Zhu, Antonio Torralba, and Wojciech Matusik. 2019. Learning the signatures of the human grasp using a scalable tactile glove. *Nature* 569, 7758 (2019), 698–702.

Paul M. Torrens and Simin Gu. 2021. Real-Time Experiential Geosimulation in Virtual Reality with Immersion-Emission. In *Proceedings of the 4th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation* (Beijing, China) *(GeoSim '21).* Association for Computing Machinery, New York, NY, USA, 19–28. https://doi.org/10.1145/3486184.3491079

Ayumu Tsuboi, Mamoru Hirota, Junki Sato, Masayuki Yokoyama, and Masao Yanagisawa. 2017. A proposal for wearable controller device and finger gesture recognition using surface electromyography. In *SIGGRAPH Asia 2017 Posters*. 1–2.

Mickeal Verschoor, Dan Casas, and Miguel A Otaduy. 2020. Tactile rendering based on skin stress optimization. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 90–1.

Christian Von Hardenberg and François Bérard. 2001. Bare-hand human-computer interaction. In *Proceedings of the 2001 workshop on Perceptive user interfaces*. 1–8.

Chengde Wan, Thomas Probst, Luc Van Gool, and Angela Yao. 2018. Dense 3d regression for hand pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5147–5156.

Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. 851–860.

Eric Whitmire, Hrvoje Benko, Christian Holz, Eyal Ofek, and Mike Sinclair. 2018. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

Francis R Willett, Donald T Avansino, Leigh R Hochberg, Jaimie M Henderson, and Krishna V Shenoy. 2021. High-performance brain-to-text communication via handwriting. *Nature* 593, 7858 (2021), 249–254.

Changcheng Wu, Qingqing Cao, Fei Fei, Dehua Yang, Baoguo Xu, Hong Zeng, and Aiguo Song. 2020. sEMG Feature Optimization Strategy for Finger Grip Force Estimation. In *International Conference on Intelligent Robotics and Applications*. Springer, 184–194.

Changcheng Wu, Qingqing Cao, Fei Fei, Dehua Yang, Baoguo Xu, Guanglie Zhang, Hong Zeng, and Aiguo Song. 2021. Optimal strategy of sEMG feature and measurement position for grasp force estimation. *PloS one* 16, 3 (2021), e0247883.

Feng Xu, Yang Zheng, and Xiaogang Hu. 2020b. Real-time finger force prediction via parallel convolutional neural networks: a preliminary study. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 3126–3129.

Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020a. Earbuddy: Enabling on-face interaction via wireless earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

Zhiqi Yin, Zeshi Yang, Michiel Van De Panne, and KangKang Yin. 2021. Discovering diverse athletic jumping strategies. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–17.

Shigeo Yoshida, Yuqian Sun, and Hideaki Kuzuoka. 2020. Pocopo: Handheld pin-based shape display for haptic rendering in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, et al. 2018. Finger-Ping: Recognizing fine-grained hand poses using active acoustic on-body sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–10.

Qin Zhang, Li Fang, Qining Zhang, and Caihua Xiong. 2022. Simultaneous estimation of joint angle and interaction force towards sEMG-driven human-robot interaction during constrained tasks. *Neurocomputing* 484 (2022), 38–45.

Yihui Zhao, Zhiqiang Zhang, Zhenhong Li, Zhixin Yang, Abbas A Dehghani-Sanij, and Shengquan Xie. 2020. An EMG-driven musculoskeletal model for estimating continuous wrist motion. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 12 (2020), 3113–3120.

Yixin Zhu, Chenfanfu Jiang, Yibiao Zhao, Demetri Terzopoulos, and Song-Chun Zhu. 2016. Inferring forces and learning human utilities from videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3823–3833.