Investigating the Impact of Deformable, Movable, and Rigid Surfaces on Force-Input Interactions

JAMES NASH, University of Bath, UK
KIM SAUVÉ, University of West England, UK
ADWAIT SHARMA, University of Bath, UK
CHRISTOPHER CLARKE, University of Bath, UK
JASON ALEXANDER, University of Bath, UK

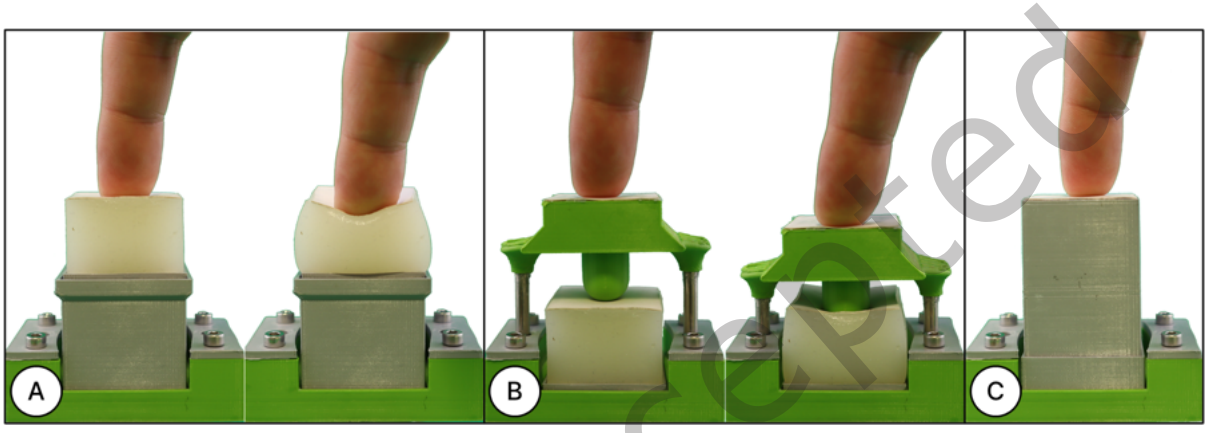

Fig. 1. We compared two compliance types: (A) *Deformable* samples with soft, elastic surfaces, and (B) *Movable* samples, in which a rigid platform can be displaced downwards, of different stiffnesses, to (C) a *Rigid* sample (C).

The force modality fundamentally transforms the interaction space of traditional touch input. When paired with compliant devices, which deform under force and provide immediate haptic feedback, there is potential to enhance user interactions significantly. However, the effects of compliance on force-input remain under-explored, with limited understanding of their full potential. This paper presents the first systematic investigation of the impact of deformable, movable, and rigid surfaces on user performance and experience through three rigorous studies (each N = 28). The results reveal previously unreported effects, including 1) higher maximum comfortable forces on deformable surfaces, 2) user preference for soft and deformable surfaces over rigid surfaces, and 3) improved ability to maintain force input on softer surfaces. These results highlight the benefits of compliant surfaces, contrasting with the dominant use of force-input on rigid devices. These findings guide researchers and designers in optimising user experience and performance of force-input interactions.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **User studies**.

Authors' addresses: James Nash, University of Bath, Bath, UK, jn688@bath.ac.uk; Kim Sauvé, University of West England, Bristol, UK, khps22@bath.ac.uk; Adwait Sharma, University of Bath, Bath, UK, as5339@bath.ac.uk; Christopher Clarke, University of Bath, Bath, UK, cjc234@bath.ac.uk; Jason Alexander, University of Bath, Bath, UK, jma73@bath.ac.uk.

ACM 1557-7325/2024/5-ART
https://doi.org/10.1145/3736409

Additional Key Words and Phrases: Force Input, Non-Rigid Interactions, Deformable, Movable, User Study, Soft Surfaces, Pressure Input

1 INTRODUCTION

Force-input can radically transform traditional binary touch interactions, enabling continuously variable, space-efficient input with a single finger. The granularity and flexibility of this additional input channel have enhanced existing interaction techniques [3, 48, 57, 81, 92] and enabled new capabilities, such as force ray-cast and force-gestures [15, 17, 31, 35, 59]. Force-input interactions have traditionally been limited to rigid devices, such as tablets [98], phones [34, 36] or trackpads [58, 60]. However, these interactions can also be performed on '*Compliant Devices*', which deform in response to user-applied force. Compliant devices represent a significant [2] and rapidly evolving [26, 32, 44, 63–65, 101, 105] field of HCI. They enable complex and novel input techniques [27, 94] and unique interaction experiences [50, 54]. Furthermore, compliant devices uniquely alter force-input interaction because the surface moves or deforms around the user's finger, creating a bi-directional interaction where the user provides force input while the device provides continuous haptic feedback. However, despite their increasing prevalence and distinct haptic qualities, our understanding of how compliance affects force-input interactions is limited.

While the distinctions between compliant and rigid devices are evident, compliant devices are not a homogenous group, and there are additional factors that affect how they respond to force-input interactions. Compliant devices have an associated *'Stiffness'*, which is the amount the device changes shape in response to an applied force, and a *'Compliance Type'*, which alters the tactile feedback [71] and the possible user inputs—for example, deformable devices enable gestures such as pinching, twisting, shearing, and bending [6]. There are two compliance types: *'Deformable'* (Figure 1A) and *'Movable'* (Figure 1B). Deformable devices [27, 41, 54, 85] are soft and elastic, deforming around the user's finger, altering the contact surface area and providing *kinesthetic* and *cutaneous* feedback [71]. In contrast, movable devices [44, 50, 69] have a rigid surface which moves along a set path, delivering only kinesthetic feedback in response to an applied force In this work, we implement the deformable interface as a silicone cube for direct user interaction. The movable interface is implemented as a rigid platform that moves in a single direction, allowing the user to indirectly compress the same silicone cube when pressing. These constructions aim to create equivalent deformable and movable devices for our experiments. In our everyday lives movable devices are normally made of rigid elements connected to springs, motors and linear actuators.

Understanding how stiffness and compliance type affect performance and user experience during force-input interactions is important for the proper design and effective use of compliant devices. There are two areas where the compliance characteristics may have particular impacts. First, Mizobuchi et al. [49] has shown that increasing the input force required reduces performance on rigid surfaces. Along with this, higher force inputs may also increase fatigue and discomfort due to the physical exertion required. However, deformable devices could alter this effect because surface deformation causes changes in the finger contact area, reducing the pressure on the user's finger. Second, the enhanced tactile feedback compliant devices provide could reduce reliance on visual feedback during force-input interactions, which typically depend on visual cues [17, 24, 68, 98]. Improving the performance of gaze-free force-input opens up new opportunities for interaction and can free up screen space by reducing the need for on-screen feedback.

Prior work has explored different compliance types individually when users are provided with visual feedback. For example, Fruchard et al. [24] examined how the stiffness of deformable samples affected force-input interactions and showed that high force levels were more demanding for soft surfaces, but that performance was similar across three different levels of stiffness. Schmitz et al. [66] investigated movable devices in the context of thumb-index pinching interactions which revealed softer counterforces reduced jitter. However, it is unclear how these different compliance types compare against each other in the context of simple force-input interactions such as pressing a button or touchscreen with the index finger. To address this, we pose three research questions

to explore how deformable, movable, and rigid (Figure 1C) devices affect force-input interactions across a range of stiffness levels:

- **RQ1** How do stiffness and compliance type affect users comfortable force-input range?
- **RQ2** How do stiffness and compliance type affect the force-input interaction experience *with* visual feedback?
- **RQ3** How do stiffness and compliance type affect the force-input interaction experience *without* visual feedback?

To address these research questions, we built a bespoke force-input device with seven interchangeable samples of varying compliance and a novel movable platform which allowed deformable samples to be used in a movable configuration. We used this apparatus to systematically evaluate how surface samples of different compliance types and stiffnesses impact user force-inputs, with and without visual feedback. Study 1 (N = 28) explored how the physical properties of an input surface impact a user's preferred force-input range (*RQ1*), with the resultant ranges used for the subsequent two studies. Study 2 (N = 28) and Study 3 (N = 28) addressed *RQ2* and *RQ3* respectively and evaluated user performance when making discrete force-input selections. In Study 2, users received visual feedback that clearly displayed their applied force, while in Study 3 participants received no visual feedback and instead had to rely on their proprioception. To support transparency and future research, the aggregated dataset from each of the three studies is publicly available at [53].

Our analysis reveals the previously unreported effects of deformable, movable, and rigid surfaces on force-input interactions. Our contributions are:

(1) A novel, bespoke Force Input Device comprising a force sensor and sample mount that allows for accurate force input on rigid, deformable, and movable samples, used across 15,288 trials over the three studies.
(2) Empirical evidence that **users apply higher maximum comfortable forces on deformable surfaces compared with rigid and movable surfaces** (Study 1). This highlights how deformability can increase the range over which a user can comfortably apply a force, and demonstrates the need to adjust the input range for force-input interactions on rigid or movable devices.
(3) Empirical evidence that there is a **significant impact of compliance on selection without visual feedback**, however **compliance does not affect selection performance with visual feedback** (Studies 2 and 3).
(4) Empirical evidence that the stiffness and compliance type significantly impact a user's ability to maintain a constant force over time irrespective of whether visual feedback is provided, with **soft samples reducing input variation the most** (Studies 2 and 3).
(5) Qualitative analysis revealing a strong subjective **preference for force-input interactions on soft and deformable surfaces, with a consistent aversion to rigid and hard surfaces** (Studies 2 and 3).
(6) A series of design recommendations specific to findings from each of the studies, paired with overarching recommendations that draw together the outcomes across all studies.

2 RELATED WORK

2.1 Compliant Interfaces

Compliance is a property of a material or object describing how it deforms when a force is applied. As described in the introduction, compliant devices can be either *movable* and made up of rigid elements that move with respect to each other, or *deformable* and made from soft, elastic materials.

Compliant interfaces are common in our everyday lives: keyboards [1], joysticks [67] and game-controllers [9] all use compliance to provide immediate tactile feedback. The stiffness of these example compliant devices is *fixed*, making them *Passive-Compliant*, however, more advanced *Active-Compliant* devices which can *dynamically* change their stiffness are becoming more common in industry (e.g. the Playstation 5 adaptive triggers [14])

and research, with devices such as surfaces [54], touchscreens [18], phones [103] and VR controllers [77] being demonstrated. The ability of active-compliant devices to dynamically change their haptic force feedback significantly increases the range of potential uses, including material and object replication [51, 69], online shopping [54], medical simulation [50] and compliant tangible notifications or emojis [103]. However, these active-compliant devices tend to require complex sensing, control and actuation systems.

As previously mentioned, compliant devices can be further categorised into *deformable* (Figure 1A and *movable* (Figure 1B). Movable devices (such as keyboards, joysticks or pin arrays) are devices made up of rigid elements which can be moved or reconfigured with respect to each other, providing kinesthetic feedback. They are common in everyday settings [1], with researchers having proposed ways of creating bespoke haptic responses [44]. Movable devices excel at rendering complex haptic force feedback [104], which can augment and enhance input interactions, however, input is typically restricted to moving an element along a 1D path. Active movable devices [50, 69, 70], can produce high-resolution haptic force feedback, which can enable encoding of information into the surface [18] and simulating of physical properties [51].

Deformable devices are "those that (1) are entirely or in part made of soft and malleable materials (e.g., rubber), (2) require physical input to be deformed, and (3) allow users to input in ways that are unlikely (if not impossible) with rigid interfaces" [6]. In contrast to movable devices, deformable devices provide both cutaneous *and* kinesthetic feedback [72]. This is due to their soft malleable nature, which also allows users to provide input by squeezing, twisting, bending and stretching [83]. This creates the opportunity for unique interactions, such as sculpting on *Trampoline* [27] or creating a deformable workspace [93]. These interactions are not possible on movable devices, however, the haptic force-feedback on deformable devices is less detailed and distinct, meaning haptic features such as clicks or bumps are not currently possible. There are also several active deformable devices [54, 102, 103], which enable these unique input techniques on a variable stiffness device. Though these interactions are not explored in this paper, they highlight the significance of compliant devices and their potential to radically transform interactions.

The significant potential and increasing prevalence of both deformable and movable devices emphasize the importance of understanding the effect of compliance on force input. Changes to the physical properties may unintentionally alter the input experiences. It may also be possible to improve force-input interaction by dynamically altering the compliance depending on the context.

2.2 Force Input Interactions

Force provides a continuous input channel for interaction, in both visual and non-visual contexts. Zooming [48, 81] and menu selection [99] provide baseline continuous 1D interactions. Force can also augment spatial input, enabling a user to change the scroll speed by varying the applied force [3], facilitating 2D navigating [15] and dynamically altering the granularity of a slider with force [57]. These approaches all provide visual feedback in some form, allowing the user to adjust their force accordingly. They also require the user to be able to precisely control the force applied and, in some cases, maintain the target force for extended periods.

Several other approaches forgo visual feedback, and utilise the force input channel to allow users to quickly select between a discrete set of levels. Commercial examples, such as Apple's "force touch" [4] allows a user to make a "force click" to select an alternative option, quickly accessing a preview or message conversation. Force gestures [31, 36] require users to be able to consistently recall and apply discrete force levels, or patterns of force, to act as a shortcut. The range of force-augmented interaction techniques highlights the ability of force to radically change the input experience. However, all of these techniques require users to be able to accurately apply and control a force input. Understanding how the input surface impacts performance can help designers create more expressive systems and maximise the input bandwidth.

2.3 Force Input Studies

As highlighted in the previous subsection, the viability of force-input interactions hinges on the accuracy and precision at which users can apply force to a device. For this reason, there have been a range of studies conducted investigating the performance of force-input interaction under different conditions. These studies primarily aim to understand the ability of users to accurately and efficiently apply a force to select a target, to aid the design of force-augmented input interactions. Table 1 shows the prior work conducted, along with Studies 2 and 3 that are presented in this paper. In this section, we discuss the key experimental elements of these studies, including *Feedback*, *Levels*, *Input Type*, *Selection Type*, *Selection method*, *Level types and Force Ranges* and *Input Surface*.

Visual Feedback: Force input studies tend to have one of two levels of visual feedback: *Full*, in which the participant sees which item they're currently selecting and; *None*, in which a participant sees no feedback before selection (Table 1, *Visual Feedback*). Ramos and Balakrishnan [59] used *Partial* visual feedback when only the target cell changed colour, when a force was applied within the target cell range. While *Reducing* visual feedback was implemented by Chen [12], where the level of feedback went from full to none while the participant maintained their applied force. However, there are currently no proposed force input techniques which use *Partial* or *Reducing* visual feedback. This explains the focus on *Full* and *None* feedback, and as such are both included within our work.

Levels: When conducting a discrete force input, the force input range will be mapped over a series of levels (Table 1, *Levels*), and applying the force within a level selects the corresponding cell (Figure 7). The visual feedback significantly impacts the number of levels (granularity) of the selection task. Using full visual feedback, Fruchard et al. [24] achieved ~95% input accuracy with 20 levels. In contrast, the accuracy of force input without visual feedback can drop below 60% on only four targets [91].

Input Type: In the majority of research to date, a user applies force directly onto the surface using their index finger (Table 1, *Input Type*). We investigate finger force input applied perpendicular to a horizontal surface, due to the prevalence of this technique. However, other research has explored different styles of force-input including Ramos et al. [58] who used force input with a pen, Schmitz et al. [66] who evaluated counterforce for pinch gestures and finally Chen [12] who considered force applied horizontal to a vertical surface.

Table 1. This table provides an overview of studies related to Force-Input selection. The studies highlighted in grey represent the two studies presented in this paper. The table includes participant numbers (N) and selection methods (QR for Quick Release). A dash indicates that data for the respective category is not available.

ID	Paper	Visual Feedback	Levels	Input Type	Selection Type	Selection Method	Level Type	Force Range	Compliance Type	Stiffness	N
1	Pressure Widgets [58]	Full	4/6/8/ 10/12	Pen	Discrete Selection	QR	Pre-Set	NR	Rigid	Rigid	12
2	Pressure Widgets [58]	Partial	4/6/8/ 10/12	Pen	Discrete Selection	QR	Pre-Set	NR	Rigid	Rigid	12
3	Presstures [60]	None	2/3	Finger	Discrete Selection	Gesture	User Set	0-1000fg	Rigid	Rigid	8
4	Exploration of pressure input [91]	Full	4/6/8/10	Finger	Discrete Selection	Dwell QR/Stroke	Pre-Set	0-4N	Rigid	Rigid	12
5	Exploration of pressure input [91]	None	2/3/4/5	Finger	Discrete Selection	Dwell QR/Stroke	Pre-Set	0-4N	Rigid	Rigid	12
6	Squish This [24]	Full	12/16/20	Finger	Discrete Selection	QR	User Set	0-5.51N (4.80, 6.57)	Deformable	Shore Hardness: 50A, 20A, 10A	24
7	Squeezy-Feely [66]	Full	25/50/75	Pinch	Discrete Selection & Hold	Clicker	Pre-Set	-	Movable	1.12 N/mm, 2.24 N/mm	18
8	Human haptic interaction with soft objects [12]	Reducing	-	Finger (Side)	Hold	-	Pre-Set	2-8N	Deformable (W/Thimble)	3.3N/mm, 2.20N/mm, 1.65N/mm	3
9	Study 2	Full	20/40/60	Finger	Discrete Selection & Hold	Clicker	Pre-Set	205-683gf	Rigid Deformable Movable	Shore Hardness: 30A, 50-00, 10-00	28
10	Study 3	None	4	Finger	Discrete Selection & Hold	Clicker	Pre-Set	205-683gf	Rigid Deformable Movable	Shore Hardness: 30A, 50-00, 10-00	28

Selection Type: Selection type (Table 1, *Selection Type*) is how the force is transformed into a system input. Most research has used a discrete selection type, where a user applies a force before making a confirmation to select a level, mirroring a traditional selection interaction. In contrast, Chen [12] conducted a continuous (hold) selection, in which the applied force is input over a time frame. Schmitz et al. [66] augmented the traditional discrete selection with a hold period after confirmation, asking participants to maintain the applied force for three seconds. This approach tested the ability of the participant to make a discrete selection, while also measuring their ability to control that force over time. In our work, we follow this approach of discrete selection and hold.

Selection Method: Research has been conducted to evaluate the impact of the selection method on discrete force selection, proposing methods such as hold/dwell [58], quick release (QR) [16] and gestures [60]. However, there has not been any research investigating the impact of compliance on the selection method, therefore it is not clear if and how this will have an impact. Schmitz et al. [66] used a clicker, held in the non-dominant hand, to try and remove any effect of selection. Furthermore, effects similar to the "Heisenberg Effect of Spatial Interaction" [100], where discrete inputs disturb the applied value, may influence force-input selection. This could be counteracted by having separate hands applying force and making selections.

Level Type and Force Range: Discrete force input requires the input levels (Table 1, *Levels*) to be mapped to a force range, with a minimum and maximum force that define the bottom and top of the levels respectively. Most papers use a set force range which is consistent for all participants and conditions (Table 1, *Force Range*). However, Fruchard et al. [24] allowed users to set their own range, over which the levels were distributed, while Rendl et al. [60] allowed users to set the individual levels (Table 1, *Level Type*). These approaches may offer some improvements, however, as with the selection method, it's not clear how force range and compliance interact with task performance, and it's plausible that changes in force range between participants could affect performance. Although using a pre-set force range will increase consistency between participants, what that range should be is not clear from the existing literature.

Surface: The surface is the physical object onto which the user applies their force input. Most studies have used rigid surfaces (Table 1, *Compliance Type*), however, some papers have considered force input on compliant surfaces. Fruchard et al. [24] using silicone cubes to create a deformable surface for force input, with their results showing a limited effect of stiffness on selection time and crossings. While Schmitz et al. [66] evaluated input performance on a movable device, using a mechanical device with springs to provide resistance when conducting a pinching interaction. The input for this device was the index finger-thumb distance and they did see a significant impact of counterforce when conducting pinch gestures when compared with no counterforce. Chen [12] used silicone rubber samples for the deformable surfaces, and a rubber cot and a rigid thimble were placed on the user's finger to reduce cutaneous feedback, going some way to replicating the differences between deformable and movable surfaces.

2.3.1 Summary. This demonstrates a need for studies evaluating the impact of compliant characteristics on force-input interactions. Compliant interfaces, both movable and deformable, offer unique interaction experiences, with active-compliant devices further enhancing these through dynamic haptic feedback. Existing research demonstrates the potential of force-input to improve current interactions and enable novel ones, but their success relies on users' ability to make accurate and precise inputs. While this has been studied on rigid surfaces, compliant surfaces remain under-explored. Understanding the effects of compliance on interaction design is essential to fully realise the potential of force-input techniques across various contexts

3 EXPERIMENTAL SETUP & APPARATUS

The same physical experiment setup was used across all three studies; we describe the setup and apparatus within this section. The two primary components of the setup are a bespoke force sensor—the Force Input Device (FID)—which sensed the participants applied force, and custom *surface samples*, which were mounted onto the FID

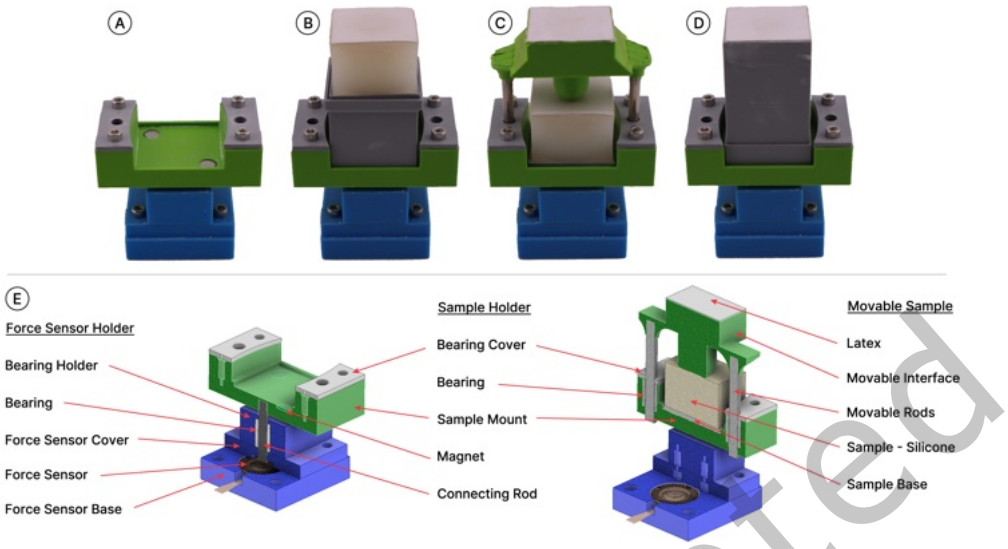

Fig. 2. Overview of the Force Input Device (FID) and its components. (A) The FID is shown in isolation. (B) The FID with a deformable sample, (C) movable sample, and (D) rigid sample attached. (E) Detailed construction of the FID, with the internal components labelled, including two cross-sectional views.

and deformed differently in response to participants' applied force. The setup additionally included a clicker to confirm selections, a display for the study GUI, and a box to cover the participant's hand for selected conditions.

3.1 Force Input Device (FID)

The FID, shown in Figure 2A, was designed as both a mount for the interchangeable surface samples and a sensor to measure the user's applied force. We chose this approach as force sensing should not impact the physical properties of the sample, therefore finger or surface-mounted sensors would be inappropriate. Previous work commonly uses force-sensing resistors (FSRs) to capture user input. However, these have several limitations, including the requirement for discrete components and circuitry to obtain a readable signal [79] along with the need for calibration and mapping, and conversion between pressure and force [89]. We identified the *TE Connectivity FX29 Compact Compression Load Cell*[1] as an alternative that addresses these challenges. They are pre-calibrated and carry out on-board conversion, transmitting the force reading via I^2C, they also provide much higher accuracy (hysteresis of 0.8% compared with 5% for an FSR [2]). The limitation of this sensor is that it does not provide a flat surface on which a sample can be mounted (as an FSR would). Instead, a housing is required which can transfer user-applied force directly onto the force sensing area.

The construction of the FID is shown in Figure 2E. At the base of the FID was the *force sensor housing*, a 3D printed PLA case to hold the FX29 Compression Load Cell, along with two linear bearings. The *sample holder* was constructed from 3D printed PLA with two 5mm steel rods extending from the base which interfaced with the bearings in the sensor housing. This held the steel rods perpendicular to the force sensor, ensuring consistent force transfer and also limiting the sample holder to one degree of freedom, sliding vertically. The sample holder had a flat square surface in the centre, with two neodymium magnets embedded into it. These mated with two

[1]TE Connectivity FX29: https://www.te.com/usa-en/product-CAT-FSE0006.html
[2]OHMITE FSR Datasheet: https://www.ohmite.com/res-fsr-rev-4.23/

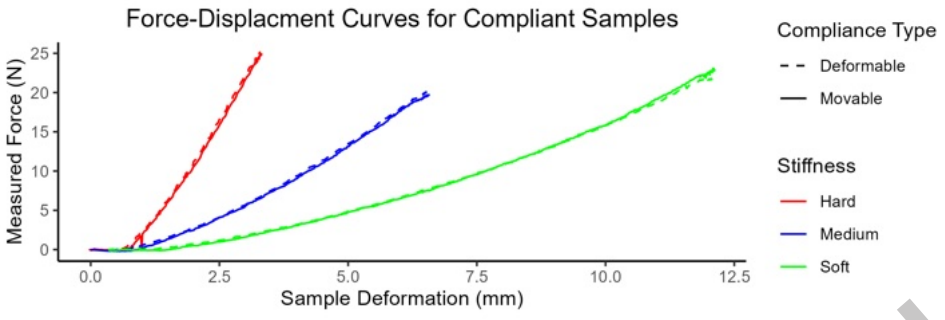

Fig. 3. Force-displacement curves for the six compliant samples from the robot arm compliance evaluation. The three different stiffness levels are represented by different colours, while the two compliance types are distinguished by dashed and solid lines.

magnets in the base of each sample to ensure the surface sample and the sample holder were firmly connected. The sample holder also had two bearings, one on either side of the sample, through which the *movable sample* could slide, providing low-friction, vertical-only movement. We share the 3D models for the FID in the supplementary material to support reproducibility.

3.2 Surface Samples

Seven different surface samples were used during the studies, all with an area of 20 mm × 20 mm. The rigid (Figure 2D) and deformable (Figure 2B) samples were each made of a single material, plastic and silicone respectively. The movable (Figure 2C) samples were mechanical, fabricated out of rigid surface (3D printed PLA) mounted on two 4 mm steel rods running through two linear bearings (Figure 2E, Movable Sample). An ellipse of size 14 mm × 20 mm [90] extended out the bottom of the PLA plate and made contact with the deformable sample placed below the plate. As the user pushed the plate down, the ellipse would press into the silicone sample and deform it, providing resistance to the user.

Rigid Sample: 3D printed on an Ultimaker S5 with PLA.
Deformable, Soft: *softer than* the human finger with shore hardness 00-10. (*Ecoflex™ 00-10*[3])
Deformable, Medium: *similar to* the human finger with shore hardness 00-50. (*Ecoflex™ 00-50*[4])
Deformable, Hard: *stiffer than* the human finger with shore hardness A-30. (*Dragon Skin™ 30*[5])
Movable, Soft: The *Deformable, Soft* sample is placed in the movable holder providing soft compliance on a movable surface.
Movable, Medium: The *Deformable, Medium* sample is placed in the movable holder providing medium compliance on a movable surface.
Movable, Hard: The *Deformable, Hard* sample is placed in the movable holder providing hard compliance on a movable surface.

Each of the six deformable and movable samples was mechanically inspected using a UR3 robotic arm, which was used due to the high position control and force measurement accuracy it provided. Figure 3 shows the six force-displacement curves, demonstrating the consistency between each of the deformable-movable sample pairs for a given compliance level. All three deformable samples were 20mm deep, though the overall travel possible

[3] *Ecoflex™ 00-10*: https://www.smooth-on.com/products/ecoflex-00-10/
[4] *Ecoflex™ 00-50*: https://www.smooth-on.com/products/ecoflex-00-50/
[5] *Dragon Skin™ 30*: https://www.smooth-on.com/products/dragon-skin-30/

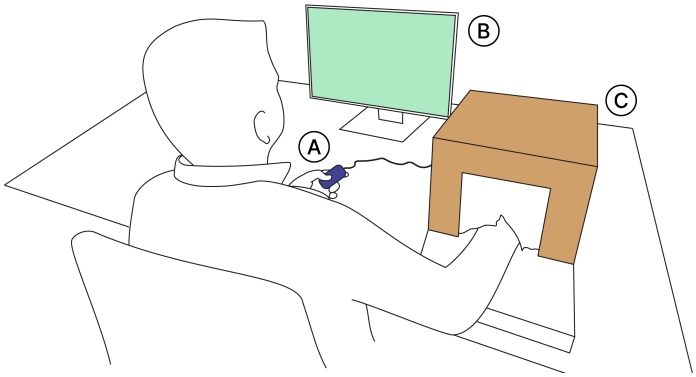

Fig. 4. Setup for Studies 1–3. Including the selection clicker (A) held in the participant's non-dominant hand; the touchscreen display (B) showing the study tasks and used input survey responses; and the opaque box (C) obscuring the participant's dominant hand and the sample from their view.

was practically constrained by the stiffness. The surface of all seven samples was covered with a 1mm sheet of latex, which ensured that other tactile cues (such as thermal conductivity or texture) were controlled.

3.3 Selection Method

During the studies, participants had to select and then hold their force for three seconds. This meant that selection methods which involve the participant lifting their hand (i.e. Quick Release) or pausing (i.e. Dwell) to select would not be suitable, as it would disrupt the three-second hold phase. We, therefore, followed the approach of Schmitz et al. [66], which also asked participants to select then hold their force, and used a clicker held in the participant's non-dominant hand, to enable this without altering their applied force.

3.4 Study Setup and Procedure

The participants sat at a desk (Figure 4), with the height adjustable chair to ensure participants were comfortable. The position of the FID could be swapped to either the right or left hand, depending on the participant's handedness. In front of the user's dominant hand is a 10 cm raised platform, providing a surface for the participant to rest their arm. The FID is sunk into the platform's surface, resulting in the top of the surface sample being in line with the platform. This allows participants to press down on it as if they were pushing into a flat display. This approach emulates using a touchscreen device, such as a phone or tablet, laid flat on a table while the user presses into it with their index finger. Although holding the device during interaction is common, this method offers two benefits. First, it limits the interaction to a single finger, simplifying the task compared to scenarios where the user holds the device. Second, it ensures a controlled and repeatable posture, reducing variation between participants. Additionally, this approach is commonly used in prior work [24, 60]. The position of the FID can be adjusted, by being moved closer or further away from the participant, depending upon their arm length. At points during the studies, the FID and the participant's dominant hand were both obscured from view, below a box. At these points, the participant placed their hand through a fabric curtain into the box and navigated to the FID. There was a physical guide under the box to ensure the hand placement was consistent. Between samples, the user could remove their hand from the box and answer the survey questions on a touchscreen display. To reduce friction on the surface and increase consistency between sample textures, participants were provided with magnesium powder to coat their index finger on their dominant hand [22]. Participants were given a wired clicker, which they held in their non-dominant hand. This clicker was used to confirm their force input during

the studies. On the desk, directly in front of the participant, was a 21-inch touchscreen display. This displayed the study software, which showed the participant the current task. This display also showed the surveys, which the participant would touch the screen to complete. During all studies, participants were greeted on arrival and provided with an information sheet, before being given the opportunity to ask any questions. They then filled out a consent form and a short survey to record age, gender, and dominant hand.

4 STUDY 1: COMFORTABLE FORCE INPUT RANGE

Ensuring users can apply force over a wide but comfortable input range could increase user preference and performance, however the relationship between comfortable force input range and the compliance of the input device is not fully understood. To select the most appropriate force input range for Studies 2 and 3, we wanted to understand how the surface sample affected user comfort during force-input interactions. Therefore, we set out to answer:

- **RQ1** How do stiffness and compliance type affect users comfortable force input range?

We achieved this by investigating the impact of surface sample on (i) the user's lowest and highest comfortable force inputs, and (ii) their ability to maintain their comfortable force inputs, reflected in the following research sub-questions:

- **RQ1.1** What is a suitable force-input range that is comfortable for users?
- **RQ1.2** How do compliance type and stiffness impact users' preferred comfortable force input range?
- **RQ1.3** How do compliance type and stiffness impact users' ability to maintain and control their preferred force inputs?

4.1 Study Design

To answer the research questions, participants were asked to input their comfortable force levels on the different SURFACE SAMPLES. Before any samples were shown, the task was contextualised for the participant. They were told to imagine they are using force to provide input into an interactive device such as a phone or tablet. They were provided with examples of setting their volume using force or scrolling through a webpage and instructed to keep in mind that their comfort levels should reflect a force they could apply for regular interactions without it becoming challenging or tiring.

The study was a within-subject design with two tasks – Task I: setting the lowest and highest comfortable force for a single selection and, Task II: setting the lowest and highest comfortable force before holding for three seconds. For both tasks, the participant applied force to the sample with their dominant hand, before using the clicker to confirm their force input. There was no lower limit on the force the participant had to apply for their force input to be registered. The study alternated between these tasks for a given sample, by showing the five following statements to the participants, instructing them to apply force in different ways.

(A) 'Please familiarise yourself with the surface and explore the range of force you feel comfortable applying, from softest to hardest'.
(B) 'Please apply your LOWEST comfortable level of force'.
(C) 'Please apply your HIGHEST comfortable level of force'.
(D) 'Please apply and HOLD your LOWEST comfortable level of force for three seconds'.
(E) 'Please apply and HOLD your HIGHEST comfortable level of force for three seconds'.

For each sample, participants were initially shown statement A, before being shown the remaining four statements (B–E) three times each (e.g. A - BCDE - BCDE - BCDE). Participants were divided into two conditions, half being asked to apply their lowest force first (B, C, D, E) and half their highest first (C, B, E, D). The study setup and procedure were followed as described in subsection 3.4. In total, the study took approximately 20 minutes.

4.1.1 Independent Variables. We manipulated the following independent variables:

- **Surface sample**: We used the seven different surface samples described in subsection 3.2.
- **Target**: Highest and lowest comfortable force.
- **Hold**: Participants were either instructed to release (Task I) or hold (Task II) the force after selection with the clicker.

When excluding the rigid sample, the remaining compliant surface samples were broken into two further independent variables:

- **Compliance Type**: Deformable and movable.
- **Stiffness**: Soft, medium and hard.

During the Study, the order of all seven surface samples was counterbalanced using a balanced Latin square to reduce any learning or fatigue effects. For each of the seven surface samples the block of four tasks (2 Target × 2 Hold) was repeated three times. This provided a total of 84 trials per participant. Across the 28 participants, this was 2352 trials in total. In total, each study took approximately 20 minutes.

4.1.2 Dependent Variables. We captured the following dependent variables during the user study, all force values are measured in grams-force (gf):

- **Applied Force** (Task I & II): The force the participant is applying when they confirm their selection (gf).
- **Hold Range** (Task II): The difference between the minimum and maximum force applied over the three-second hold phase after the participant makes their selection (gf).
- **Hold Drift** (Task II): The difference between the applied force level at the beginning and end of the three-second hold phase, measured in gf/s.
- **Hold standard deviation** (Task II): The standard deviation from the mean over the three-second hold phase (gf).

4.1.3 Participants. We recruited 28 participants (18 identified as male, 10 as female), aged between 18 to 42 (Mean = 23.6, σ = 5.6). All had full mobility of their index finger on their dominant hand, 26 were right-handed and 2 left-handed.

4.1.4 Analysis. We analysed the data to understand the impact of the dependent variables, using the following approach:

- **Applied Force (Task I & II)**: The lowest, highest and range for comfortable values were calculated and a two-way repeated measures ANOVA (RM-ANOVA), with factors of surface sample and hold was conducted on each (lowest, highest and range).
- **Hold Control (Task II)**: A two-way RM-ANOVA with factors of surface sample and target was conducted for the hold range, drift and standard deviation.

To explore the effect of stiffness and compliance type we then conducted two further three-way RM-ANOVA's, with the rigid sample removed from the data set and the remaining six samples decomposed into the two factors. For brevity, we only report effects or interactions of stiffness and compliance type for the three-way RM-ANOVA.

We log-transformed positively skewed data following the approach by Dragicevic [21]. Due to the nature of the study, outliers were possible from unintended inputs, such as accidental clicks or releasing the applied force by mistake. Furthermore, there was no upper limit on the possible applied force, and no upper or lower limit on any of the hold variables, which increased the potential for outliers. Winzorisation has been shown as a robust method for reducing the effect of outliers on the standard deviation of normally distributed data, without significantly impacting it [20]. We used graphical inspection of the data combined with our knowledge of the variables to apply appropriate Winsorizations, as recommended by Ghosh and Vogt [25]. During Study

Table 2. Significant results of the RM-ANOVA, for Study 1. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	F	DFn	DFd	p	η^2
1	Highest Comfortable Force	Sample	2.636	4.14	111.91	0.036*	0.089
2	Highest Comfortable Force	Hold	4.257	1	27	0.049*	0.136
3	Lowest Comfortable Force	Sample	3.06	3.81	102.88	0.022*	0.102
4	Lowest Comfortable Force	Hold	5.009	1	27	0.034*	0.156
5	Lowest Comfortable Force	Compliance Type	12.04	1	27	0.002**	0.31
6	Applied Force Range	Compliance Type	9.48	1	27	0.005**	0.26
7	Hold Range	Target	117.38	1	27	< .001***	0.81
8	Hold Drift	Target	27.81	1	27	< .001***	0.51
9	Hold Standard Deviation	Target	116.35	1	27	< .001***	0.81

1, all the dependent variables had a similar likelihood of outliers therefore outliers were limited using a 95% Winsorization across all trials. The assumption of normality was tested using a Shapiro-wilks test and Mauchly's test of sphericity was used to check the assumption of sphericity, where it indicated the assumption was violated, we used Greenhouse-Geisser to correct. We report the effect size of a factor with the partial eta square (η_p^2), this is qualified in terms of small (< .01), medium (< .06) and large (< .14) based off Cohen [13]. The results of the RM-ANOVA were reported only when there was a significant interaction, with a summary of them provided in Table 2, however the full tables of results for all studies are included in the supplementary material. When the RM-ANOVA showed a significant effect for a given factor ($p < 0.05$), a Holm-Bonferroni corrected pairwise t-test was used for post hoc-analysis. Due to the positively skewed force ranges, we report the median ($\bar{X}$) along with the 95% median confidence intervals, unless stated otherwise.

4.2 Results

4.2.1 Comfortable Force Ranges. There was a median lowest comfortable value of $\bar{X}$ = 205 gf [5%, 95%CI = 196 gf, 216 gf] and the highest comfortable value of 683 gf [650 gf, 729 gf], with a comfortable force range of 439 gf [410 gf, 467 gf]. These median lowest and highest comfortable values were used as the comfortable force range

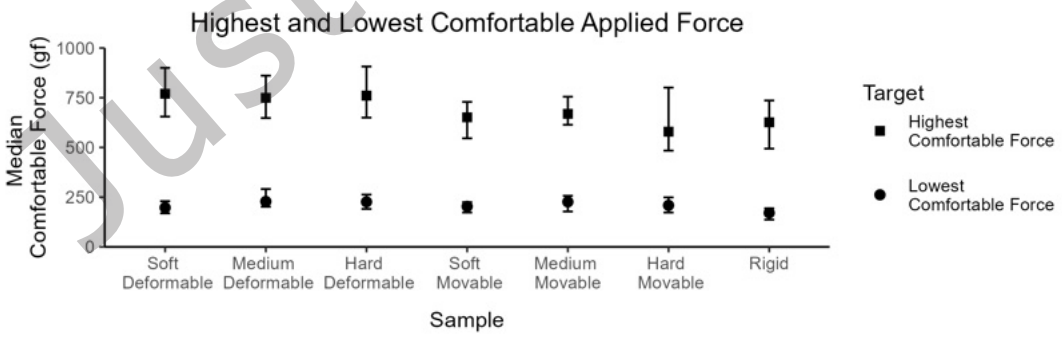

Fig. 5. Median lowest and highest comfortable forces for each of the seven Surface samples, for Study 1. The highest comfortable force is notably higher for the three deformable samples compared to the three movable and rigid samples. The graph includes 95% upper and lower median confidence intervals.

for Studies 2 and 3. The results for the lowest and highest comfortable force values for all samples are shown in Figure 5.

There was a significant effect ($F_{4.14,111.91} = 2.64, p = .036, \eta_p^2 = .089$) of SAMPLE on **lowest comfortable force value**, with a medium effect size. The post hoc tests only showed a significant difference between hard deformable and rigid. There was also a significant main effect of HOLD ($F_{1,27} = 4.257, p = .049, \eta_p^2 = .136$) on lowest comfortable force value, with a medium effect size. The lowest comfortable force was higher for the hold task ($\bar{X}$ = 209 gf [192 gf, 225 gf]) compared to no hold ($\bar{X}$ = 204 gf [191 gf, 216 gf]). There was no significant effect of either STIFFNESS or COMPLIANCE TYPE on the lowest comfortable force.

There was a significant effect ($F_{3.06,102.88} = 306, p = .022, \eta_p^2 = .102$) of SAMPLE on **highest comfortable force value**, with a medium effect size. The post hoc tests did not show any significant difference between samples. There was a significant main effect of HOLD ($F_{1,27} = 5.01, p = .034, \eta_p^2 = .156$) on the highest comfortable force value, with a large effect size. With the highest comfortable force also higher on hold ($\bar{X}$ = 693 gf [648 gf, 756 gf]) compared with no hold ($\bar{X}$ = 672 gf [629 gf, 733 gf]). The three-way RM-ANOVA showed the effect of COMPLIANCE TYPE was significant ($F_{1,27} = 12.04, p = .002, \eta_p^2 = .31$) with a large effect size. The highest force applied to deformable samples ($\bar{X}$ = 758 gf [697 gf, 838 gf]) was larger than that applied on the movable samples ($\bar{X}$ = 647 gf [586 gf, 689 gf]), indicating that participants found it more comfortable to apply large forces to deformable surfaces.

There was no significant effect of either SAMPLE or HOLD on **comfortable force range**. The three-way RM-ANOVA showed that COMPLIANCE TYPE had a significant effect on comfortable force range ($F_{1,27} = 9.48, p = .005, \eta_p^2 = .26$) with a large effect size. As with the highest comfortable force, participants applied a larger range to deformable samples ($\bar{X}$ = 493 gf [435 gf, 551 gf]) compared with movable samples ($\bar{X}$ = 408 gf [343 gf, 446 gf]).

Hold Measures: The hold measures reported below are calculated during Task II when the participant presses the clicker and lasts for three seconds afterwards, during which the participant is instructed to maintain their force. The results for the hold range, drift and standard deviation for the lowest and highest comfortable force values are shown in Figure 6.

4.2.2 Hold Range. Over all trials, the median range of force values during the hold was 69 gf [63 gf, 76 gf]. The analysis showed that only the effect of TARGET on hold range was significant ($F_{1,27} = 117.38, p < .001, \eta_p^2 = .81$), with a large effect size. The post hoc tests showed a significant difference ($p < .001$) between the range of force values for the lowest comfortable force ($\bar{X}$ = 40 gf [37 gf, 44 gf]) compared with the range for the highest

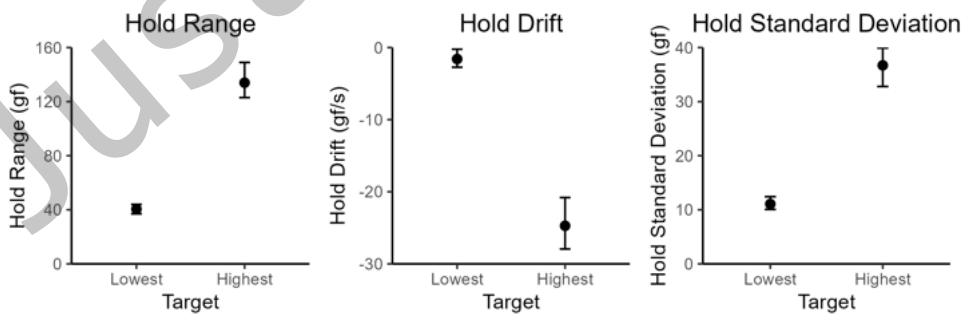

Fig. 6. Median hold range, drift, and standard deviation between participants' lowest and highest comfortable forces. The impact of TARGET on the magnitude of all three measures can be seen, as all are larger when applying the highest comfortable force. The graph includes 95% upper and lower median confidence intervals.

comfortable force ($\bar{X}$ = 134 gf [123 gf, 149 gf]). There was no significant effect of SAMPLE, COMPLIANCE TYPE or STIFFNESS on hold range.

4.2.3 Hold Drift. We calculated the amount of drift during the three-second hold by carrying out a linear regression and obtaining its slope, reported in gf/s. A positive hold drift means the applied force increased after the three-second hold, while a negative hold drift means it decreased. The *hold drift* measure helps us understand if the input force is undergoing a consistent change in either direction. The median hold drift over all trials was −8.4 gf/s [−10.0 gf/s, −6.9 gf/s], suggesting that force decreases over time during the hold. The analysis showed that only the effect of TARGET on hold drift was significant ($F_{1,27} = 27.81, p < .001, \eta_p^2 = .51$), with a large effect size. The post hoc tests showed participants tended to be stable at the lowest comfortable force value ($\bar{X}$ = −1.6 gf [−2.7 gf, −0.2 gf]), while their force reduced when holding the highest comfortable force value ($\bar{X}$ = −24.7 gf/s [−27.9 gf/s, −20.8 gf/s]). There was no significant effect of either SAMPLE, STIFFNESS or COMPLIANCE TYPE on hold drift.

4.2.4 Hold Standard Deviation. Finally, we report the standard deviation of the force values during the three-second hold which characterises the total movement around the mean. The median hold standard deviation was 19.2 gf [17.7 gf, 21.3 gf]. The analysis showed that the effect of TARGET on hold standard deviation was significant ($F_{1,27} = 116.35, p < .001, \eta_p^2 = .81$), with a large effect size. The post hoc tests showed a significant difference ($p < .001$) in standard deviation between lowest comfortable force value ($\bar{X}$ = 11.1 gf [10.1 gf, 12.4 gf]) and highest comfortable force value ($\bar{X}$ = 36.7 gf [32.8 gf, 39.9 gf]). There was no significant effect of either SAMPLE, STIFFNESS or COMPLIANCE TYPE on hold standard deviation.

4.3 Summary and Discussion

This study set out to answer **RQ1** which we broke down into **RQ1.1–1.3**. Our results demonstrate that a suitable comfortable force-input range lies between 205 gf [5%, 95%CI = 196 gf, 216 gf] and 683 gf [650 gf, 729 gf] (RQ1.1). They also provide insights into how the surface sample has an impact on the lowest comfortable force and that users prefer to apply higher comfortable forces and larger comfortable force ranges on deformable surfaces (*RQ1.2*). The study did not show a significant impact of COMPLIANCE TYPE on hold standard deviation, hold range or hold drift (*RQ1.3*). However, there was a large and significant difference for hold standard deviation, hold range and hold drift between the lowest and highest comfortable force ranges.

4.3.1 Lowest and Highest Comfortable Input Force. The lowest and highest comfortable input force commonly used in prior studies are presented by Fruchard et al. [24]; they use a force range of $0 - 5.51$N and pointed to other works with ranges of $0 - 4$ N [49], $0 - 10$ N [46] and $0 - 3.5$ N [98]. This study has established a median highest comfortable force value of 682 gf (6.69 N) and a median lowest comfortable force value of 205 gf (2.01 N). These results contrast with the prior work, with the highest input force being higher than most prior studies, indicating the range over which users can apply comfortable forces is not being fully utilised. Along with this, all prior work used an input force of 0 N for the lowest level, while our results also show that users do not necessarily find it comfortable to apply such a low level of force. The use of a threshold for the lowest force input may improve the user experience. Although our range is for the most common type of touch-input interactions, the input posture can significantly impact the force users apply to devices [82], therefore it's likely that the comfortable force ranges will change depending on posture.

From these results we also see significant variations across participants, highlighted in the large confidence intervals. This indicates that comfortable force is a personal metric and depends heavily on the individual. Interestingly, the variation is larger for the highest comfortable force value compared with the lowest comfortable force value, even as a percentage. Exactly what traits affect preference for the highest comfortable force value is unclear, but factors could include fingertip surface area, fingertip sensitivity, and finger or arm strength.

Furthermore, the results from these studies will likely be specific to the demographics and may be skewed because of the age (Mean = 23.6, σ = 5.6) or gender (18 identified as male, 10 as female) distributions of the participants. Our results provide clear evidence that users should be able to set their own force-input range, especially the highest force value.

4.3.2 Impact of Sample. Our results reveal a significant difference in the highest comfortable force value set on deformable vs. movable surfaces, showing that deformable samples allow users to apply higher comfortable force values before reaching discomfort. As we controlled for the level of hardness, which was validated with a robotic arm (see Figure 3), we can speculate this difference may be caused by the increased contact area on deformable surfaces as they are compressed, reducing the pressure on the fingertip. This provides a clear example of how the different compliance types, which have been used interchangeably, can distinctly alter the force input experience and therefore must be treated differently. Researchers and practitioners should be mindful that the highest comfortable force values set on deformable surfaces will likely be uncomfortable on movable surfaces.

Our results for the effect of STIFFNESS differ from those of Fruchard et al. [24] despite some overlap in the sample stiffness ranges, with medium (Shore Hardness 00-50) and hard (Shore Hardness A-30) samples in this study being comparable to soft (Shore Hardness A-10) and medium (Shore Hardness A-20) samples in their paper. Fruchard et al. [24] found higher comfortable forces for their softer samples compared to the medium and hard samples, whereas our study shows no significant effect of stiffness on the highest comfortable force levels. There are several potential explanations for this. First, the most likely explanation is that the FSR used to measure force in their study was placed between the participant's finger and the surface. This would affect the surface deformation, and therefore the cutaneous feedback, because the FSR is flexible but not elastic. Second, we used a clicker to confirm selection whereas Fruchard et al. [24] used the "Quick Release" (QR) technique in which participants quickly remove the applied force to make a selection. It could be that on softer samples, the deformation from the applied force and the resultant finger movement alters the QR gesture, or that participants apply more force to perform the gesture more easily. Finally, it could be that the relationship between comfort and stiffness is non-linear, and that our samples reflect where this relationship plateaus.

4.3.3 Hold Phase. Our results show that it is much harder to hold higher force levels consistently (Figure 6), as can be seen by significantly higher ranges and standard deviations of applied force during the highest comfortable force hold. Along with this, participants drifted downwards much more when applying their highest comfortable forces, providing further evidence that they find it difficult to maintain these forces for a duration of time. The study did not find any impact of SAMPLE on the hold performance, however the study showed that participants were applying higher forces on the deformable samples which could have contributed to increased variation. Therefore, we will consider the relationship between COMPLIANCE TYPE and hold control in studies 2 and 3.

Recommendations. Based on our results, we draw the following **two recommendations** for setting force ranges for force selection applications:

(1) *The force input range should not by default begin at zero; instead, it should be defined by user input.*
(2) *Users should be able to set their own highest input force value to account for large variations between users.*

4.3.4 Limitations. There are several limitations to the results of Study 1. First, although the participants were provided with some context before completing the task, there was no context within the application and it was an abstract task. It may be possible that a user's preferred range varies depending on the context or application. Next, no longitudinal data was collected, instead all the tasks were completed during a single session. It may also be the case that user preferences change over time, as they become more familiar with force input interactions. In addition, a comfortable force is a subjective concept that may vary significantly between individuals. A more objective measure could be the highest and lowest force a user can apply without significant fluctuations. Finally,

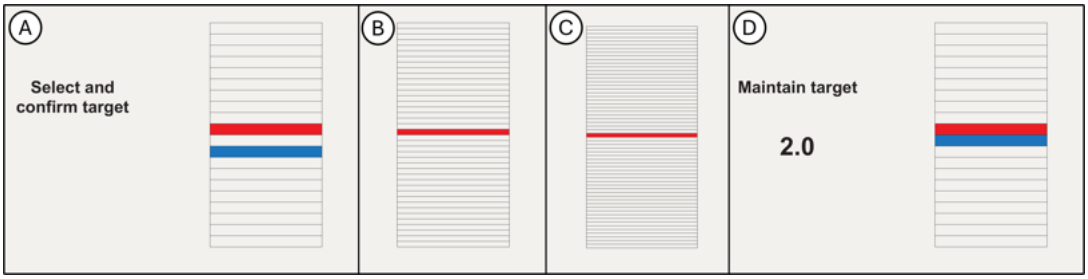

Fig. 7. The interface for Study 2, shows 20-cell scale (A), 40-cell scale (B) and 60-cell scale (C). The target cell is highlighted in red and applied force is in blue. (D) shows the interface during the hold phase, after selection.

we only investigated one type of input posture, and prior research has shown that different force input ranges exist for other postures such as thumb-finger pinching [66].

5 STUDY 2: DISCRETE FORCE SELECTION WITH VISUAL FEEDBACK

Study 1 established the preferred lowest and highest comfortable force values, along with the impact of the compliance type when applying the highest comfortable force. These findings are used in this Study, as well as Study 3, which both seek to investigate discrete force selection on compliant surfaces. The use of force for discrete selection is well established, with various examples in the literature [58, 60]. However, as highlighted in the introduction, the physical characteristics of a surface can potentially influence force input [12, 24], something which we sought to understand with the following research question:

- **RQ2** How do stiffness and compliance type affect the force input interaction experience *with* visual feedback?

We build on prior research [12, 24, 33], expanding in three key ways. First, we consider (i) deformable surfaces, (ii) movable surfaces and (iii) a rigid surface together, which has not been done before (see Table 1). Second, we use empirically derived pre-set values for the lowest and highest force levels to control for the effect of input force on user performance. Third, we conduct a discrete selection and hold task, evaluating users' ability to locate and maintain a target force. To answer *RQ2*, we broke it down into the following research sub-questions:

- **RQ2.1** How do the input SURFACE SAMPLE, TARGET, and SCALE affect the users' accuracy and efficiency of force selection with visual feedback?
- **RQ2.2** How do the input SURFACE SAMPLE, TARGET, and SCALE affect the user's ability to maintain a constant force input after selection, with visual feedback?
- **RQ2.3** How do the input SURFACE SAMPLE and SCALE impact the demand, workload, and user experience of a force selection task with visual feedback?

5.1 Study Design

We used a within-participants repeated measures design and, following the approaches of similar studies [24, 66], we conducted a visual target acquisition and selection task. Participants interacted with the system by applying force onto the FID described in subsection 3.1: increasing the force raises the selection value, and decreasing the force lowers it. As the participants changed their applied force, the selected cell moved (shown by the blue bar in Figure 7A). The scale is mapped linearly over the force range, which was set based on the medium comfortable values calculated from Study 1. The lower comfortable value of 205 gf was used as the bottom of the first cell, with the higher comfortable value of 683 gf as the top of the last cell. This mapping was consistent across all trials.

Following the approach of Schmitz et al. [66] the task required the participant to navigate to the target value, select it using the clicker and hold the force for three seconds. The participant began the task by resting the index finger of their dominant hand on the surface without applying force. They were then shown a target value (red) on the linear scale and had to apply force onto the surface, shown by the force cell (blue), to navigate to the target cell before using a clicker in their non-dominant hand to select their input. Participants were instructed to reach the target as quickly and accurately as possible. They were told that if they found the target was impossible to reach they should get as close as they could and confirm. After the selection, the visual feedback was maintained and the participant had to hold their final force for three seconds (Figure 7D), after which they were prompted to release their force. Once they release their force, there is a short pause before they are shown the next target value. During all tasks, the participant's dominant hand and the sample are hidden from view to eliminate potential visual feedback from their finger or the surface, ensuring that the only feedback comes from the on-screen visuals.

Once all the trials for a given combination of SURFACE SAMPLE and SCALE were completed the participant filled out the NASA TLX questionnaire, and after all SCALES for a given SURFACE SAMPLE the participants filled out a Likert scale questionnaire and were encouraged to take a short break. Once all SURFACE SAMPLES had been seen the participants completed a short semi-structured interview. The study setup and procedure were followed as described in subsection 3.4. In total, each study took approximately one hour, with most participants taking two or three breaks.

5.1.1 Independent Variables. The following independent variables were manipulated:

- **SURFACE SAMPLE**: We conducted trials with the seven different surface samples described in subsection 3.2.
- **SCALE**: We varied the granularity of the interaction scale, by changing the total number of cells between 20 (Figure 7A), 40 (Figure 7B), and 60 (Figure 7C). We selected these values based on pilot testing, in which these scales showed significant variance in ease of selection.
- **TARGET**: There are five different target levels (Low: 10%, Mid-Low: 30%, Mid: 50%, Mid-High: 70%, and High: 90%), from which participants were shown one per trial. For all scales, the target cell was the one which occupied the given percentage level (where 0% is the lowest force level and 100% is the highest).

By excluding the rigid sample, the remaining compliant SURFACE SAMPLES could be broken into the following two further independent variables:

- **COMPLIANCE TYPE**: There are two compliant types; deformable and movable.
- **STIFFNESS**: There are three levels of stiffness tested; soft, medium and hard.

For each of the seven SURFACE SAMPLES, participants would perform 30 trials, consisting of five TARGETS repeated twice, on each of the three SCALES. The order of the SURFACE SAMPLES was counter-balanced using a balanced Latin square. For each SURFACE SAMPLES, the SCALE were shown in order of increasing difficulty (20, 40, and 60), while the order of TARGETS was randomised for each repetition. This resulted in 7 SURFACE SAMPLES × 3 SCALES × 5 TARGETS × 2 repetitions = 210 trials per participant. This resulted in a total of 5880 trials over all 28 participants.

5.1.2 Dependent Variables. We captured the following dependent variables during the user study.

- **Success**: Whether the participant was on the target cell when they made their selection.
- **Task Completion Time**: The time taken by the participant, between applying the minimum threshold force and pressing the clicker to select the input (s).
- **Crossings**: The number of times the participant crossed into the target cell. We do not include leaving the cell as a crossing.
- **Hold Range**: The difference between the maximum and minimum force applied over the three seconds after the participant makes their selection (gf).

- **Hold Drift**: The drift in the force level applied over the three seconds after selection, measured in gf/s.
- **Hold Standard Deviation**: The average deviation from the mean over the three seconds after selection (gf).

To evaluate the impact of the surface on the task, participants were asked to conduct the NASA TLX (Appendix A.1) after completing each combination of scale and sample (21 questionnaires per participant). After completing the 30 trials within a given sample, they then completed a short questionnaire consisting of four 7-point Likert scale questions asking how pleasant, comfortable, stiff, and deformable they perceived the surface (Appendix A.2). Finally, after completing all trials the participants completed a Sample Perception Survey (Appendix A.3) – a post-experiment semi-structured interview where the participants ranked the seven samples in order of their preference for using them during the study. The participants were also asked about their perceptions and preferences of the stiffness and compliance styles. The semi-structured interviews were audio-recorded with participants' consent and the recordings were transcribed.

5.1.3 Participants. We recruited 28 participants (16 identified as male, 12 identified as female), aged between 18 to 38 (Mean = 21.0, σ = 4.3). All had full mobility of their index finger on their dominant hands, with 25 right-handed and 3 left-handed. They were compensated with £10 each for taking part in the study.

5.1.4 Analysis. Success is a dichotomous dependent variable and it was not possible to analyse using an RM-ANOVA. Instead, we used a binomial logistic generalised linear mixed model (optimised with BOBYQA), with SURFACE SAMPLE, SCALE and TARGET. We also created a four-way model with the rigid sample removed from the data set and the remaining six samples decomposed into STIFFNESS and COMPLIANCE TYPE – the remaining two factors were SCALE and TARGET. Following an approach by Nakagawa and Schielzeth [52] we calculate the explanatory power of the models for both fixed and random effects and for just the fixed effects. We calculated the chi-square (χ^2) and the p-values of the individual effects using a likelihood ratio test (LRT). We used an approach proposed by Stoffel et al. [80] to calculate the semi-partial R^2 (R^2_{sp}) for each effect, which describes the level of variance explained by the effect. Finally, we conduct pairwise comparisons of the log-odds between levels of fixed effects with significant differences. We then report the estimated marginal means (μ) and the 95% confidence intervals. The fixed effects were coded as ordinal variables.

All dependent variables apart from success were evaluated using a three-way RM-ANOVA with SURFACE SAMPLE, SCALE and TARGET as factors to compare compliant versus rigid samples. We also then conducted a four-way RM-ANOVA with the rigid sample removed and six other samples decomposed into STIFFNESS and COMPLIANCE TYPE. For brevity, we only report effects or interactions of STIFFNESS and COMPLIANCE TYPE for the four-way RM-ANOVA and four-way generalised linear mixed model.

For the RM-ANOVA, dependent variables which were positively skewed were log-transformed to correct for this. As with Study 1, we used Winzorastion to reduce the effect of outliers. However, outliers for certain dependent variables were less common. First, the number of crossings could not increase significantly by accident (i.e. by pressing the clicker), therefore outliers were less common and it was Winzorised to 1% lower and 99% upper. Next, although the completion time has a lower limit of zero, there is a risk of outliers at the higher end from input errors. Therefore, the completion time was Winsorized to 5% lower and 95% upper. Outliers were more common for all three hold variables, as they could change significantly through a mistaken input (i.e. releasing the force) and were possible at both the lower and upper ends. Therefore a 5% lower and 95% upper winsoriations was applied consistently across the three variables. The assumption of normality was tested using a Shapiro-Wilks test and Mauchly's test of sphericity was used to check the assumption of sphericity, where it indicated the assumption was violated, we used Greenhouse-Geisser to correct. We report the effect size of a factor with the partial eta square (η^2_p), this is qualified in terms of small (< .01), medium (< .06) and large (< .14) based off Cohen [13]. The results of the RM-ANOVA were reported only when there was a significant interaction,

Table 3. Significant results of the binomial regression and RM-ANOVA, for Study 2. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2 / F	DFn	DFd	p	R^2_{sp} / η^2
1	Success	Scale	163.34	4	108	< .001***	0.0296
2	Success	Target	360.91	2	54	< .001***	0.0641
3	Success	Scale × Target	23.07	8	216	0.003**	0.0072
4	Completion Time	Scale	104.83	1.53	41.29	< .001***	0.80
5	Completion Time	Target	147.99	2.25	60.81	< .001***	0.85
6	Completion Time	Scale × Target	2.84	8.00	216.00	0.005 *	0.10
7	Crossings	Scale	120.03	1.25	33.78	< .001***	0.82
8	Crossings	Target	6.90	2.23	60.08	0.001 **	0.20
9	Crossings	Scale × Target	2.26	5.46	147.34	0.046 *	0.08
10	Hold Range	Sample	5.65	6.0	162.0	< .001***	0.17
11	Hold Range	Scale	123.20	2.0	54.0	< .001***	0.82
12	Hold Range	Target	89.98	2.2	58.7	< .001***	0.77
13	Hold Range	Scale × Target	2.60	8.0	216.0	0.01 *	0.09
14	Hold Range	Stiffness	9.77	2.00	54.00	< .001***	0.27
15	Hold Drift	Scale	3.89	2.00	54.00	< .001***	0.13
16	Hold Drift	Target	74.73	2.58	69.77	< .001***	0.74
17	Hold Drift	Scale × Target	6.90	4.36	117.85	< .001***	0.20
18	Hold Standard Deviation	Sample	6.01	6.00	162.00	< .001***	0.18
19	Hold Standard Deviation	Scale	157.30	2.00	54.00	< .001***	0.85
20	Hold Standard Deviation	Target	81.50	2.13	57.40	< .001***	0.75
21	Hold Standard Deviation	Scale × Target	2.50	8.00	216.00	0.013 *	0.09
22	Hold Standard Deviation	Stiffness	10.32	2.00	54.00	< .001***	0.28

with a summary of them provided in Table 3, however the full tables of results for all studies are included in the supplementary material. When the RM-ANOVA showed a significant effect for a given factor ($p < 0.05$), a Holm-Bonferroni corrected pairwise t-test was used for post hoc-analysis. We either report the mean (μ) along with the 95% confidence intervals, or the median ($\tilde{X}$) along with the 95% median confidence intervals.

Responses from the subjective questionnaires were analysed using Friedman tests with the "coin" package in R [39] to check for effects within each factor in non-parametric data before conducting a post hoc pairwise Wilcoxon signed-ranked test. This approach did not let us detect interactions. We report the median score for the questionnaires along with the Mean Absolute Deviation (MAD). Finally, qualitative data was collected from the semi-structured interviews, transcripts were reviewed, and 94 quotes were extracted. These were categorised using inductive thematic analysis [7, 8] to iteratively craft themes that comprised participants' reflections.

5.2 Results

5.2.1 Success. The overall success across all trials was $\mu = 67\%([60\%, 73\%])$, this ranged from the low-target on the 20-level scale ($\mu = 89\%[84\%, 92\%]$) to the highest target on 60-level ($\mu = 37\%[29\%, 46\%]$). The success results for each SAMPLE are shown in Figure 8, *success*, and for each combination of SCALE and TARGET in Figure 9, *success*. The three-way binomial logistic generalised linear mixed model was created with SAMPLE, TARGET and SCALE as fixed effects with an interaction between TARGET and SCALE and participant number as a random effect. This model yielded the lowest Akaike information criterion (AIC) value when comparing models. The explanatory

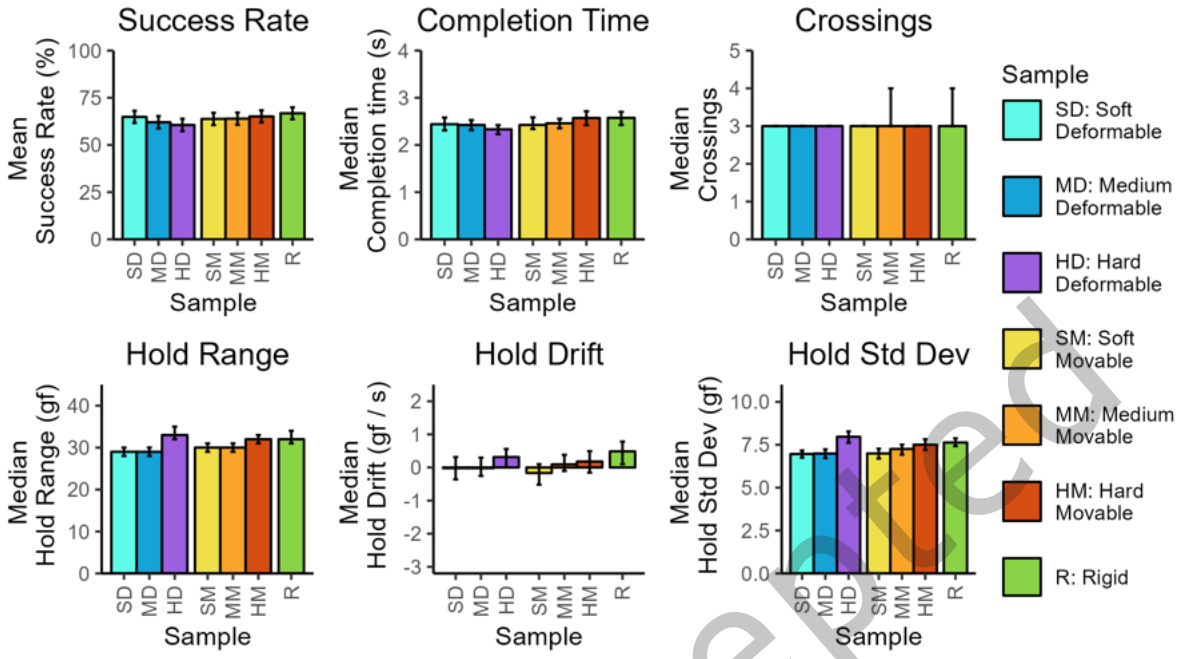

Fig. 8. Median values for (left to right from top left) success rate, completion time, crossings, hold range, hold drift and hold standard deviation across all seven SAMPLES in Study 2. The graph includes 95% upper and lower median confidence intervals.

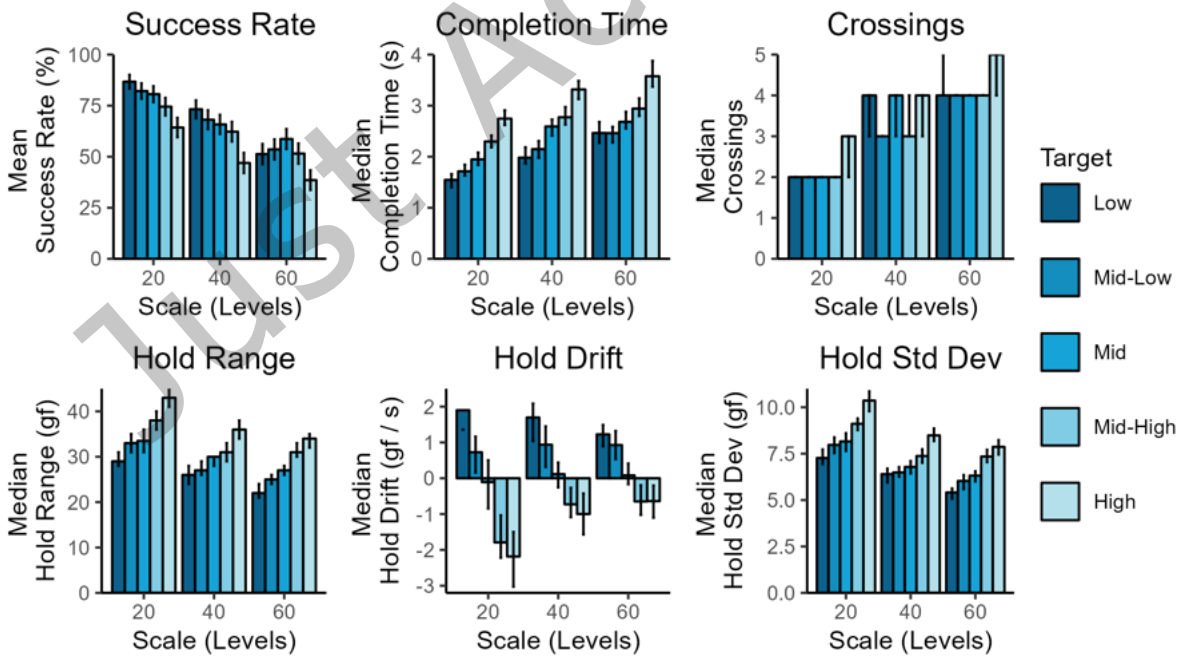

Fig. 9. Median values for (left to right from top left) success rate, completion time, crossings, hold range, hold drift and hold standard deviation across each combination of TARGET and SCALE in Study 2. The graph includes 95% upper and lower median confidence intervals.

power of the model was ($Cond.R^2 = .256$) for both fixed and random effects and ($Marginal.R^2 = .117$) for just the fixed effects. There was a significant ($\chi^2_{8,216} = 23.07, p = .003, R^2_{sp} = .0072$) two-way interaction effect between TARGET and SCALE on success. There was a simple main effect ($p < .001$) of TARGET on all levels of SCALE, with performance best on 20-level ($\mu = 81\%[76\%, 86\%]$) and worst on 60-level ($\mu = 51\%[43\%, 59\%]$). Post hoc tests showed how performance on the high target was significantly ($p < .001$) worse compared with all other targets. Along with this, on the 20-level and 40-level scales, the success rate falls as the TARGET increases. However, this trend is not seen for the 60-level scale, and instead, success is highest on the mid-target. A four-way model was also created with fixed effects of STIFFNESS, COMPLIANCE TYPE, TARGET and SCALE, with an interaction between TARGET and SCALE, and participant number as a random effect, its explanatory power was ($Cond.R^2 = .258$) for both fixed and random effects and ($Marginal.R^2 = .115$) for fixed effects. The four-way model did not show significance for COMPLIANCE TYPE or STIFFNESS.

5.2.2 Completion Time. Across all trials the median completion time was 2.45 s [2.41 s, 2.50 s]. The completion time results for SAMPLE are shown in Figure 8, *completion time*, and for SCALE and TARGET in Figure 9, *completion time*. The three-way RM-ANOVA showed a significant ($F_{8,216} = 2.84, p = .005, \eta^2_p = .1$) two-way interaction between SCALE and TARGET with a medium effect size. There was a simple main effect of target ($p < .001$) for all levels of scale, with completion time increasing with SCALE from 2.07 s [1.99 s, 2.13 s] for 20-level to 2.84 ms [2.75 s, 2.95 s] at 60-level. The completion time increased with the target level, with a significant difference between each target for the 20-level and 40-level scales, while the low and mid-low targets were consistent for the 60-level scale. However, the completion time was the largest on the high target for all scales. The four-way RM-ANOVA did not show significance for COMPLIANCE TYPE or STIFFNESS.

5.2.3 Crossings. Across all trials the median number of crossings was 3.0 [3.0, 3.0], the median crossings for each SAMPLE are shown in Figure 8, *crossing*, and for each combination of SCALE and TARGET in Figure 9, *crossing*. The three-way RM-ANOVA revealed a two-way interaction between TARGET and SCALE ($F_{5.46,147.34} = 2.26, p = .046, \eta^2_p = .08$), with a medium effect size. There were simple main effects of TARGET for all levels of SCALE, with the median number of crossings increasing with scale, from 2.0 [2.0, 2.0] for 20-levels to 4.0 [4.0, 4.0] for 60-levels. There was also a two-way interaction between STIFFNESS and SCALE ($F_{4,108} = 2.55, p = .043, \eta^2_p = .09$), with a medium effect size, however there were no simple main effects of STIFFNESS for any level of SCALE. The post hoc tests showed that crossings were largest on the high target, while crossings on the mid-target changed depending on the scale. The four-way RM-ANOVA did not show significance for COMPLIANCE TYPE or STIFFNESS.

5.2.4 Hold Range. Across all trials, the median hold range was 31 gf [30 gf, 31 gf]. The hold range results for each SAMPLE are shown in Figure 8, *hold range*, and for each combination of SCALE and TARGET in Figure 9, *hold range*. Example force-input graphs for the three-second hold period are shown in Figure 10, for different level targets and different scale granularities. There was a significant effect ($F_{6,162} = 5.65, p < .001, \eta^2_p = .17$) of SAMPLE on *hold range*, with a large effect size. Post hoc tests showed that the hold range for soft and medium deformable samples was significantly lower than hard deformable ($p \leq .046$) and rigid ($p = 0.046$), and that the soft movable was significantly lower than the hard deformable ($p = .014$). There was also a significant two-way interaction ($F_{8,216} = 2.6, p = .01, \eta^2_p = .09$) between SCALE and TARGET on *hold range*, with a medium effect size. There was a simple main effect of the TARGET at every level of SCALE, with the hold range decreasing from the 20-level scale ($\bar{X} = 35$ gf [34 gf, 36 gf]) to the 60-level scale ($\bar{X} = 27$ gf [27 gf, 28 gf]). The hold range is different between targets for the three scales, however they all followed the same trend which is that the hold range increases with the target (i.e. the high target exhibited the largest hold range). Along with this, there was a significant effect ($F_{2,54} = 9.77, p < .001, \eta^2_p = .27$) of STIFFNESS on *hold range*, with a large effect size. The hold range was significantly larger for the hardest ($\bar{X} = 33$ gf [32 gf, 34 gf]) compared with both the softest ($\bar{X} = 29$ gf

[28 gf, 30 gf]) and medium samples ($\bar{X}$ = 30 gf [29 gf, 31 gf]). There was no significant difference between the soft and medium samples.

5.2.5 Hold Drift. A positive hold drift shows that the applied force increases over the three-second hold phase, while a negative value means the applied force decreases. The median hold drift across the study was 0.1 gf/s [0.0 gf/s, 0.2 gf/s], with the median hold drift on each SAMPLE shown in Figure 8, *hold drift*, and for each combination of SCALE and TARGET are in Figure 9, *hold drift*. There was a significant two-way interaction ($F_{4.36,117.85} = 6.90, p <$ 001, $\eta_p^2 = .20$) between SCALE and TARGET on hold drift, with a large effect size. There was a simple main effect of TARGET on all levels of SCALE, however, there were only simple main effects of SCALE on the mid-high and high targets. The applied force increased the most on the low target ($\bar{X}$ = 1.5 gf/s [1.3 gf/s, 1.8 gf/s]) and decreased the most on the highest target ($\bar{X} = -1.2$ gf/s [−1.5 gf/s, −0.8 gf/s]), with negative drift on the high and mid-high targets reducing from 20-level to 60-level scales. There was no effect of STIFFNESS or COMPLIANCE TYPE on hold drift.

5.2.6 Hold Standard Deviation. The median hold standard deviation was 7.3 gf [7.2 gf, 7.4 gf]. The hold standard deviation results for SAMPLE are shown in Figure 8, *hold standard deviation*, and for each combination of SCALE and TARGET in Figure 9, *hold standard deviation*. There was a significant effect of SAMPLE ($F_{6,162} = 6.01, p <$ 001, $\eta_p^2 = .18$) on hold standard deviation, with a large effect size. Post hoc tests showed standard deviation on the rigid and hard deformable samples was significantly higher than the soft deformable ($p \leq .049$) and medium deformable ($p = 0.45$), and hard deformable was significantly higher than soft movable ($p = .004$). There was also a significant two-way interaction ($F_{8,216} = 2.50, p = 0.013, \eta_p^2 = .09$) between SCALE and TARGET on hold standard deviation, with a medium effect size. There was a simple main effect of the TARGET at every level of SCALE, with standard deviation decreasing from the 20-level scale ($\bar{X}$ = 8.5 gf [8.4 gf, 8.8 gf]) to the 60-level scale ($\bar{X}$ = 6.5 gf

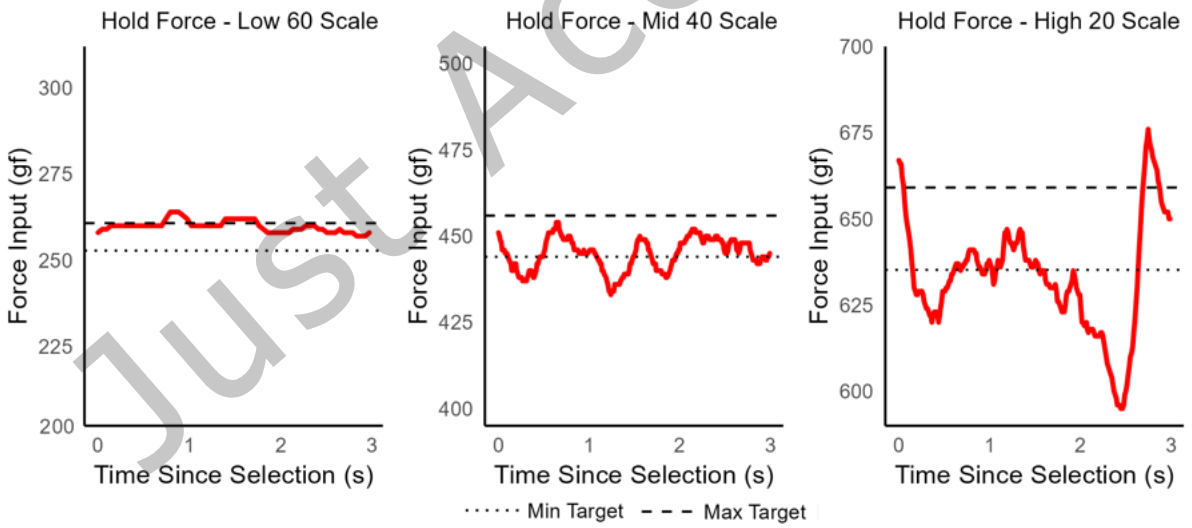

Fig. 10. Example graphs showing the applied force during the three-second hold period, for (from left to right) the lowest target on the 60 cell scale (P25), middle target for 40 cell scale (P17) and the highest target on the 20 cell scale (P6). The graphs provide an example of how hold range and hold standard deviation increase with scale size and target force. They also demonstrate the lower level of drift on the 60-cell scales and at the mid-target.

[6.3 gf, 6.6 gf]). Last, there was a significant main effect of STIFFNESS ($F_{2,54} = 10 - .32, p < 001, \eta_p^2 = .28$) on hold standard deviation, with a large effect size. Post hoc tests showed standard deviation increased with STIFFNESS, with significant differences between the hard samples ($\bar{X} = 7.7$ gf [7.5 gf, 8.0 gf]) and both the soft ($\bar{X} = 7.0$ gf [6.8 gf, 7.2 gf]) and medium samples ($\bar{X} = 7.1$ gf [6.9 gf, 7.3 gf]) at $p \leq 0.004$.

5.2.7 Demand and Workload Survey. The NASA-TLX questions were used to judge the demand and workload of discrete force selections on different scales and different samples (see Table 7). The results of the analysis of the demand and workload survey are shown in Table 5 and Table 6. For all measures of the NASA-TLX, scores increased with SCALES, with significant differences between all scales ($p \leq 0.001$), clearly indicating the increase in demand and workload as the scale increases. For SAMPLE, participants reported differences in physical demand, temporal demand and performance, however, there were no significant effects seen between individual samples in the post hoc tests. When evaluating the effects of COMPLIANCE TYPE and STIFFNESS on demand and workload, the only significant effect ($p = .004$) was of COMPLIANCE TYPE on physical demand, with physical demand being significantly lower on the deformable samples.

5.2.8 Sample Perception Survey. The results for participants' perception of the comfort and pleasantness of the seven samples are shown in Figure 11, while the results of the analysis are provided in Table 8 and Table 9. Participants answered on a Likert scale from 1 (lowest) to 7 (highest).

The analysis showed a significant effect of SAMPLE ($p \leq 0.001$) and STIFFNESS ($p \leq 0.009$) on both measures. The soft samples ($\bar{X} = 4$ [MAD = 1.48]) were rated significantly ($p = 0.007$) more pleasant than the hard samples ($\bar{X} = 3$ [MAD = 1.48]), and the hard samples ($\bar{X} = 3$ [MAD = 1.48]) was significantly less comfortable ($p \leq 0.032$) than both the soft ($\bar{X} = 4$ [MAD = 1.48]) and medium samples ($\bar{X} = 4$ [MAD = 1.48]). There was also a significant effect ($p = 0.009$) of COMPLIANCE TYPE on comfort. The deformable samples ($\bar{X} = 4$ [MAD = 1.48]) were perceived as significantly more comfortable compared with the movable samples ($\bar{X} = 3$ [MAD = 1.48]).

Participants were also asked about their perception of the physical properties of the sample, the results of which are shown in Table 10. On the whole, participants' perceptions aligned with the physical properties of the samples, however, participants perceived the deformable samples ($\bar{X} = 2$ [MAD = 1.48]) as softer than movable samples ($\bar{X} = 4$ [MAD = 2.22]).

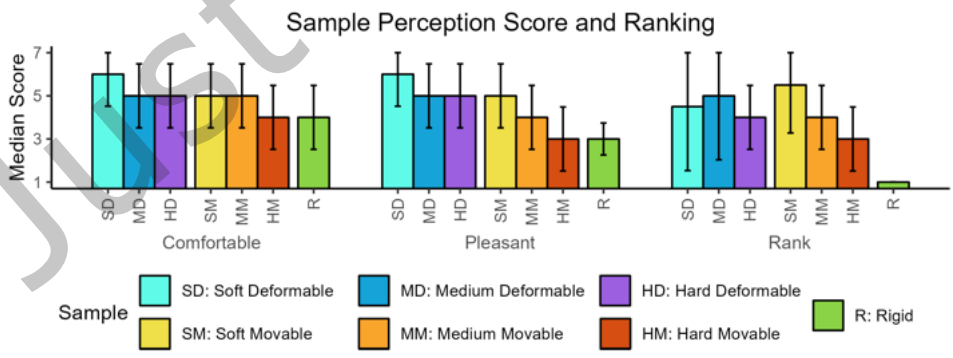

Fig. 11. Median scores for participant perceptions of how comfortable and pleasant each SAMPLE was, along with their ranking of each individual SAMPLE, for Study 2. The graph includes Mean Absolute Deviation (MAD) from the median score. Scores were given between 1 (lowest) and 7 (highest).

5.2.9 Individual Sample Ranking. During the post-study interview, participants ranked the seven samples, from one (least favourite) to seven (most favourite) in terms of their preference for using them during the study. The rankings are shown in Figure 11 and the results of the Friedman test are presented in Table 8 and Table 9. Overall, rigid was the least preferred sample, with compliance significantly impacting participants' ranking, with softer and medium samples being ranked higher.

5.2.10 Post-Study Interview. We carried out a thematic analysis on the quotes extracted from the semi-structured interviews conducted with participants after the studies, we have broken up the analysis into three sections **Perceived Performance**, **Perceived Feeling**, and **Sample Perception** based on the themes discussed by the participants.

Perceived Performance. More than half of the participants (n = 16) commented on whether a specific compliance type or stiffness made the study easier or harder (e.g. more difficult to deform, easier to complete task). The soft (5 easier/3 more difficult), medium (3/0), deformable (3/0) and movable (6/3) all leaned towards making the task easier, while hard (1/4) and rigid (1/4) were perceived as making it more difficult. Over half of the participants (n = 17) also commented on whether a specific compliance type or stiffness impacted their performance in terms of "control" or "accuracy". The control and accuracy of the soft (5 improved/3 reduced), medium (3/0), and deformable (4/2) samples were somewhat positively viewed, while the results for movable (6/1) were most clear, with participants commenting on the increased control it provides. For instance, P5 stated *"I'd say this one [Movable], allows you to push down and hold it there. I feel like it was easier to control than any of these."* Twelve participants provided additional reasons why samples impacted their performance. Most common among these was how samples improved their awareness of applied force, in particular for soft (3) and movable (3) samples, with P20 saying about soft *"I kind of know how much pressure I'm applying. It's easier to gauge how much pressure"*. Participants also described how the effort was reduced (soft:1, movable:1) or increased (hard:1, rigid:2).

Participants' performance perception of different stiffness samples matches the ranking survey results. Rigid is consistently perceived as reducing performance by lowering control and accuracy. In contrast, soft samples elicit mixed opinions: some participants believe that increased travel enhances performance, while others disagree. The medium sample, though less frequently mentioned, is often viewed positively, suggesting it may offer a good compromise with its reduced travel. Regarding compliance type, perceptions of movable and deformable were mixed, but movable was more frequently viewed in a positive light.

Perceived Feeling. Participants (n = 17) expressed their preference for the sample based on how it felt to use. Eight participants mentioned how *"comfortable"* a sample was, with soft (7) and deformable (4) being mentioned the most often; while six described samples as uncomfortable, focusing mainly on hard (3) and rigid (2). Participants also described the hard and rigid samples as *"fatiguing"* (Hard:2, Rigid:1) as well as *"painful"* (H:1, R:3). These illustrate that participants perceived force-input interactions on the hard and rigid samples as negative, calling them *"uncomfortable"* or *"painful"*, and repeatedly describing the soft and deformable samples as *"comfortable"*.

Sample Perception. Five participants mentioned the amount of travel or motion on the soft samples, with two being positive towards the motion, such as P4, *"travel made it feel like the zone I was trying to hold it in was quite a lot bigger"*; while the other three disliked it, including P6: *"like you're pressing but [there's] almost nothing"*. Three participants liked soft (1) and medium (2) movable samples as they were familiar, comparing them to a piano or keyboard. Finally, several participants discussed conflicts in their preferences, usually between perceived feel and performance, such as P17: *"I'd actually say that despite this being less comfortable [Hard]. I felt like it was easier, with the precision"*; and P10 stating: *"I feel the softer ones are more comfortable, but it's harder to get exactly at the right place"*.

Summary. Overall, participants felt the samples impacted their performance, with the hard and rigid samples being seen as making the task more difficult and reducing performance, while the soft and medium samples were more divided. Softer and deformable samples were viewed as comfortable while the rigid and hard samples were painful and fatiguing, demonstrating a consistent impact of sample on perceived feeling. This demonstrates a consistent preference towards soft and deformable samples and away from hard and rigid samples for conducting force interactions.

5.3 Summary and Discussion

Study 2 was conducted to answer **RQ2**, which was divided into research sub-questions **RQ2.1–2.3**. Our results show that there is no significant difference of SAMPLE on selection performance (*RQ2.1*), however it has demonstrated how increases in TARGET detriment success, completion time and crossings, especially at the highest target level. We also see a reduction in performance across all measures with increasing SCALE. Study 2 demonstrated, for the first time, that with visual feedback the choice of SAMPLE can significantly reduce the applied range and standard deviation of force applied during hold interactions (*RQ2.2*), with the soft deformable sample having the best hold performance. The range and standard deviation when holding force-input with visual feedback both increased with TARGET, matching the trend from selection, while participants had positive drift when applying lower forces and negative drift on higher forces. All the measures of hold performance improved with increases in scale, likely due to the more granular feedback. Finally, the impact of SAMPLE on the user experience (*RQ2.3*) was significant. Although the effect of SAMPLE on workload and demand was limited, both comfort and pleasantness were higher on soft samples and lower on hard and rigid samples. This trend was also observed in both the individual sample rankings, in which rigid was consistently worst, and in the qualitative feedback in which perceptions of performance and feeling were better for soft samples.

5.3.1 Sample Impact on Hold. This is the first study to establish the impact of STIFFNESS on performance when holding force-input for short durations. Our findings reveal a significant reduction in hold range and standard deviation on softer surfaces, suggesting that they facilitate more precise hold input interactions. This observation is particularly noteworthy as it underscores the potential for optimising hold selection mechanisms in interactive systems. We speculate that softer samples may dampen small movements or vibrations of a user's finger during the hold phase, which would normally be transferred to an applied force on rigid samples. We therefore believe it's possible to leverage this dampening effect to enhance the stability and accuracy of force-based interactions.

5.3.2 Sample Impact on Preference. Along with the impact of the SAMPLE on hold performance, Study 2 also established clear user preferences for input surface. The most substantial result is the *significant dislike for conducting force input interactions on rigid surfaces*, seen consistently across the majority of participants. This finding is particularly striking, as the vast majority of force input interactions use rigid surfaces [15, 17, 48, 57]. Participants comment on how conducting force input on a rigid or hard surface is *"uncomfortable"*, *"fatiguing"* and *"painful"*, yet this is the default approach. These results strongly suggest we should move away from force input on rigid surfaces and look to provide soft and even deformable interfaces where possible to improve user force input experience. If this is not possible, designers must be aware that force input interactions on rigid surfaces will likely become tiring and unpleasant and must reduce the required force input or the time conducting the input accordingly.

5.3.3 Impact of Target on Performance. The impact of TARGET on performance is shown in Figure 9, where performance is worse on the highest target across all measures. Performance normally reduces as the target level, and therefore force required, increases. However, this is not always the trend and we observed that for success rates on the 60-level scale and crossings across all scales, performance was best on the mid-target and reduced towards the lowest and highest targets. This indicates that above a certain force level, accuracy, efficiency and

control begin to break down. It seems likely that force input interactions should be restricted to a force range below this breakdown limit. Interestingly this limit is lower than the median higher comfortable force for this set-up established in Study 1. Our results suggest that the force input range should be limited to 80% of the comfortable force range. Restricting force input interactions to this range, would increase success, and reduce completion time, crossings and hold performance.

Recommendations. Based on the findings from Study 2, we present the following **three recommendations** for force input interactions with visual feedback:

(1) *Force-hold interactions should be conducted on soft deformable surfaces, to optimise performance and user experience.*
(2) *Force-input interactions should be avoided on rigid surfaces as participants find them uncomfortable and unpleasant.*
(3) *The force range of force input interactions should be limited to 80% of the maximum comfortable force, to prevent performance breakdown observed at higher target forces.*

5.3.4 Limitations. There are several limitations to this study. The first is that within the study each participant carried out 210 force input trials, spending nearly an hour pressing with the same finger. It's unlikely a user would conduct this interaction repeatedly for this amount of time and so will have potentially increased fatigue and strained the user's finger. This may have caused participants to mention phases such as *"tiring"*, *"painful"* and *"uncomfortable"* more often, however as the samples were counterbalanced it should have affected all samples equally. It is also worth noting that the inclusion of the hold phase likely increased the physical workload, potentially contributing to overall fatigue. Excluding the hold phase, as in other studies [24], might have mitigated this effect. Another limitation of the study is that the force-input range is fixed for all samples to prevent force range from acting as a confounding factor, and we observed large variations in comfortable force-input ranges in Study 1. This meant some participants were applying forces above or below their comfortable range, which may have reduced their overall performance. It therefore may have been possible to increase the average performance if users had used personalised force ranges. Similarly, force input was mapped linearly to the applied force, although non-linear mappings, such as increasing the target width at higher forces or using "Fisheye levels" [68], could have improved performance. It is possible that such mapping techniques could provide overall improvements or make the user more consistent across the force scale.

6 STUDY 3: DISCRETE FORCE SELECTION WITHOUT VISUAL FEEDBACK

In this study, we examine discrete force selection without visual feedback. Our over-arching hypothesis is that the kinesthetic feedback that a user receives from applying a force to a compliant object improves their ability to recall the discrete force levels. Whether or not compliance impacts input during gaze-free force-input interactions has not been addressed in the literature, and we aim to address this by answering RQ3:

- **RQ3** How do stiffness and compliance type affect the force-input interaction experience *without* visual feedback?

To do so, we conduct a study which investigates how *surface sample* and the visibility of the participant's finger (*Finger visibility*) affect performance. More specifically, we break RQ3 down into the following three research sub-questions:

- **RQ3.1** How does the SURFACE SAMPLE, FINGER VISIBILITY and input TARGET affect the users' accuracy and efficiency of force selection without visual feedback?
- **RQ3.2** How does SURFACE SAMPLE, FINGER VISIBILITY, and input TARGET affect a user's ability to maintain a constant force input after selection without visual feedback?

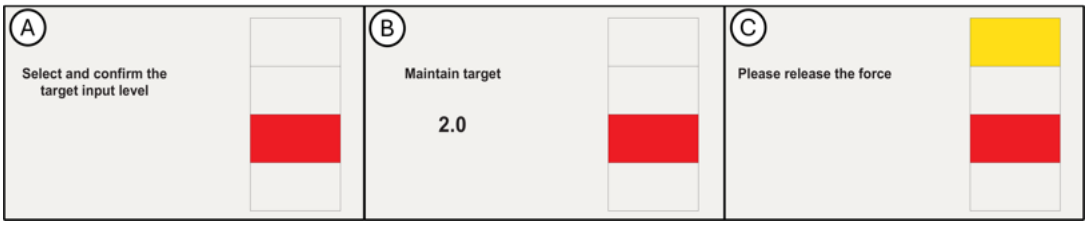

Fig. 12. Study 3 UI. The three screens show the UI during the main study for the *Selection* (A). *Hold* (B) and *Confirmation feedback* (C) stages. The red cell is the target, while the yellow cell is the feedback, indicating the cell the participant selected.

- **RQ3.3** How does the SURFACE SAMPLE and FINGER VISIBILITY impact the demand, workload, and user experience of a force selection without visual feedback?

6.1 Study Design

Building on the approaches of Wang et al. [91] and Fruchard et al. [24] we conducted a within-participant target acquisition and selection task without visual feedback. The core input interaction was identical to Study 2, with the participant applying force to the FID (subsection 3.1) to select a cell, before pressing a clicker with their non-dominant hand to confirm the selection. However, in contrast with Study 2, the participant did not receive visual feedback about which cell they were selecting before the confirmation, instead having to rely on their proprioception and recall to select the correct level.

For each sample, the participant would first go through a set training phase (described in subsubsection 6.1.1) before beginning the recall phase of the study. The recall phase involved two blocks of 12 trials each, with the participant's finger *seen* and *unseen* to understand whether or not the visual feedback from the fingertip when applying force (e.g. visibly depressing and changing colour) would affect input performance. The order of these blocks was counterbalanced for each participant and was consistent across all of their samples.

The participants were shown (Figure 12 A) a target value (red) on the linear scale and were required to apply force onto the surface to navigate to it before using a clicker in their non-dominant hand to confirm their input then hold the force for three seconds (Figure 12 B). Participants were instructed to reach the target as quickly and accurately as possible. After the selection and hold, a yellow cell showed which level they had selected when they pressed the clicker, and they were prompted to release their force (Figure 12 C). Once they release their force, there is a short pause before they are shown the next target value. For all trials, the scale had four levels and they were mapped linearly over the force range defined in Study 1, however, the top and bottom cells were unbounded (described in subsubsection 6.1.2).

After all the trials for a given block, the participant completed a short NASA TLX questionnaire covering the demand and workload during that block. Then, once both blocks for a given sample were complete, a Likert scale questionnaire was completed on their perceptions of the sample they had just been using. Participants were then encouraged to take a short break. Once all surface samples were tested the participants completed a short semi-structured interview. The study setup and procedure were followed as described in subsection 3.4, with the only difference being that the box was removed during the finger *seen* condition. Each study approximately took one hour, and participants normally took two or three breaks.

6.1.1 Training Phase. The training phase was conducted each time the participant was presented with a new sample and consists of the following three stages with gradually decreasing levels of visual feedback (see Figure 13):

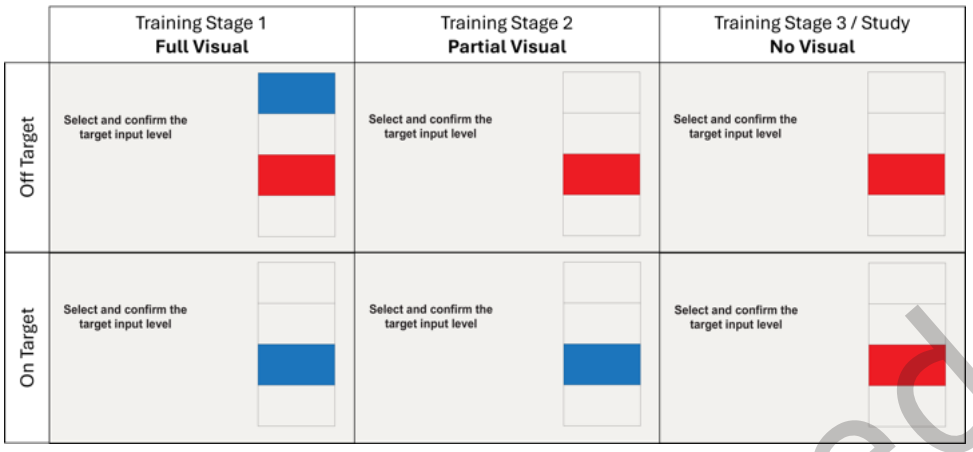

Fig. 13. The UI for training during Study 3. The three stages *Full Visual*, *Partial Visual* and *No Visual* are shown for both the incorrect force being applied (Off Target) and the correct force being applied (On Target).

(1) Full Visual: The current cell selected by the participant is highlighted in blue to inform them of their applied force at all times (Figure 13, *Full Visual*).
(2) Partial Visual: The applied force is only shown when they apply a force within the target range and the target cell changes colour from red to blue (*Partial Visual*).
(3) No Visual: Their applied force is not indicated at any point during the selection phase, however, they are still shown their selected level after selection (Figure 13, *No Visual*).

During all stages of training, participants were shown each target at least once, repeating the target a maximum of three times or until they selected it correctly. After completing all four targets within a given training stage, they progressed to the next stage. Once stage three was complete, there was a short pause (five seconds) before participants began the recall phase.

6.1.2 Scale Selection. The scale comprised four levels centred at 12.5% (low), 37.5% (mid-low), 62.5% (mid-high), and 87.5% (high), each with a width of 25% of the scale. The total width of the scale was based on the comfortable ranges established in Study 1 (the same as Study 2) with a minimum force of 205 gf and a maximum force of 683 gf. The lowest (level 1) and highest level (level 4) are *unbounded*, meaning that values below 0% (205 gf) and above 100% (683 gf) will be registered as selections for the lowest and highest levels respectively. Choosing unbounded upper and lower levels significantly reduces the difficulty of selecting these extremes. We adopted this approach for two reasons: first, it aligns with previous key work [58, 60, 91], and second, it enhances ecological validity, mirroring how force selection would likely be implemented in real-world applications. A minimum threshold of 50 gf was used to prevent a participant from selecting a level without touching the sample.

6.1.3 Independent Variables. The following independent variables were manipulated:

- **Surface sample**: We conducted trials with the seven different surface samples described in subsection 3.2.
- **Target:** We used four scale levels as described in subsubsection 6.1.2.
- **Finger visibility:** The participant's dominant hand used for input is either visible to them (*seen*) or obscured from view (*unseen*).

When excluding the rigid sample the remaining compliant SURFACE SAMPLES are broken down into the following two independent variables:

- **COMPLIANCE TYPE**: There are two compliant types; deformable and movable.
- **STIFFNESS**: There are three levels of stiffness tested; soft, medium and hard.

The three independent variables were varied in a repeated measures design. For each of the 14 combinations of SURFACE SAMPLES and FINGER VISIBILITY, participants would perform 12 trials, consisting of 3 repetitions of 4 TARGETS. The seven SURFACE SAMPLES were counterbalanced using a balanced Latin square, resulting in 14 distinct orders. Each of the 14 orders was shown with both finger-seen and finger-unseen first, resulting in a total of 28 unique orders of the SURFACE SAMPLES and FINGER VISIBILITY, one being seen by each participant. This resulted in 7 SURFACE SAMPLES × 4 TARGETS × 2 FINGER VISIBILITY × 3 (repetitions) = 168 trials per participant. Along with this, participants conducted at least 12 (3 × 4 TARGETS) training repetitions for each sample, for a total of 84 training trials per participant. This gave at least 252 trials in total for each participant, for a total of 7056 trials over all 28 participants.

6.1.4 Dependent Variables. We captured the following dependent variables during the user study:

- **Success**: Whether the participant was on the target cell when they made their selection.
- **Task Completion Time**: The time taken by the participant, between applying the minimum threshold force to pressing the clicker to select the input (s).
- **Hold Range**: The difference between the maximum and minimum force applied over the three seconds after the participant makes their selection (gf).
- **Hold Drift**: The drift in the force level applied over the three seconds after selection, measured in gf/s.
- **Hold Standard Deviation**: The average deviation from the mean over the three seconds after selection (gf).

We did not measure crossings in Study 3 because participants may not pass through the target area at all (if their selection ability is poor) without visual feedback and with significantly fewer levels.

Participants completed the same questionnaires and interviews as in Study 2. This involved a NASA TLX (Appendix A.1) after completing each combination of scale and sample (21 questionnaires per participant). They also answered four 7-point Likert scale questions regarding how pleasant, comfortable, and compliant the sample was, along with the softness of the surface (Appendix A.2). Finally, after completing all the trials, a semi-structured interview in which the participant ranked all the samples and discussed their preferences for sample stiffness and surface rigidity was completed (Appendix A.3).

6.1.5 Participants. We recruited 28 participants (15 identified as male, 12 identified as female, and 1 preferred not to say), aged between 18 to 58 (μ = 27.8, σ = 12.0). All had full mobility of their index finger on their dominant hands with 27 right-handed and 1 left-handed. They were compensated with £10 each for taking part in the study.

6.1.6 Analysis. We used the same analysis approach for Study 3 that was used in Study 2, described in subsubsection 5.1.4. The same bounds for Winsorization as described in Study 2 were used in Study 3. A summary of the significant results is provided in Table 4.

6.2 Results

6.2.1 Success. The overall success across all trials was $\mu = 65\%([64\%, 66\%])$. The success results for each combination of SAMPLE and TARGET are shown in Figure 14, *success rate*. The binomial logistic generalised linear mixed model was created with SAMPLE and TARGET as main effects, FINGER VISIBILITY was excluded from the model as it reduced model performance and participant number was a random effect. This model yielded a lower AIC value when comparing models with different interactions. The explanatory power of the model was ($Cond.R^2$ = .383) for

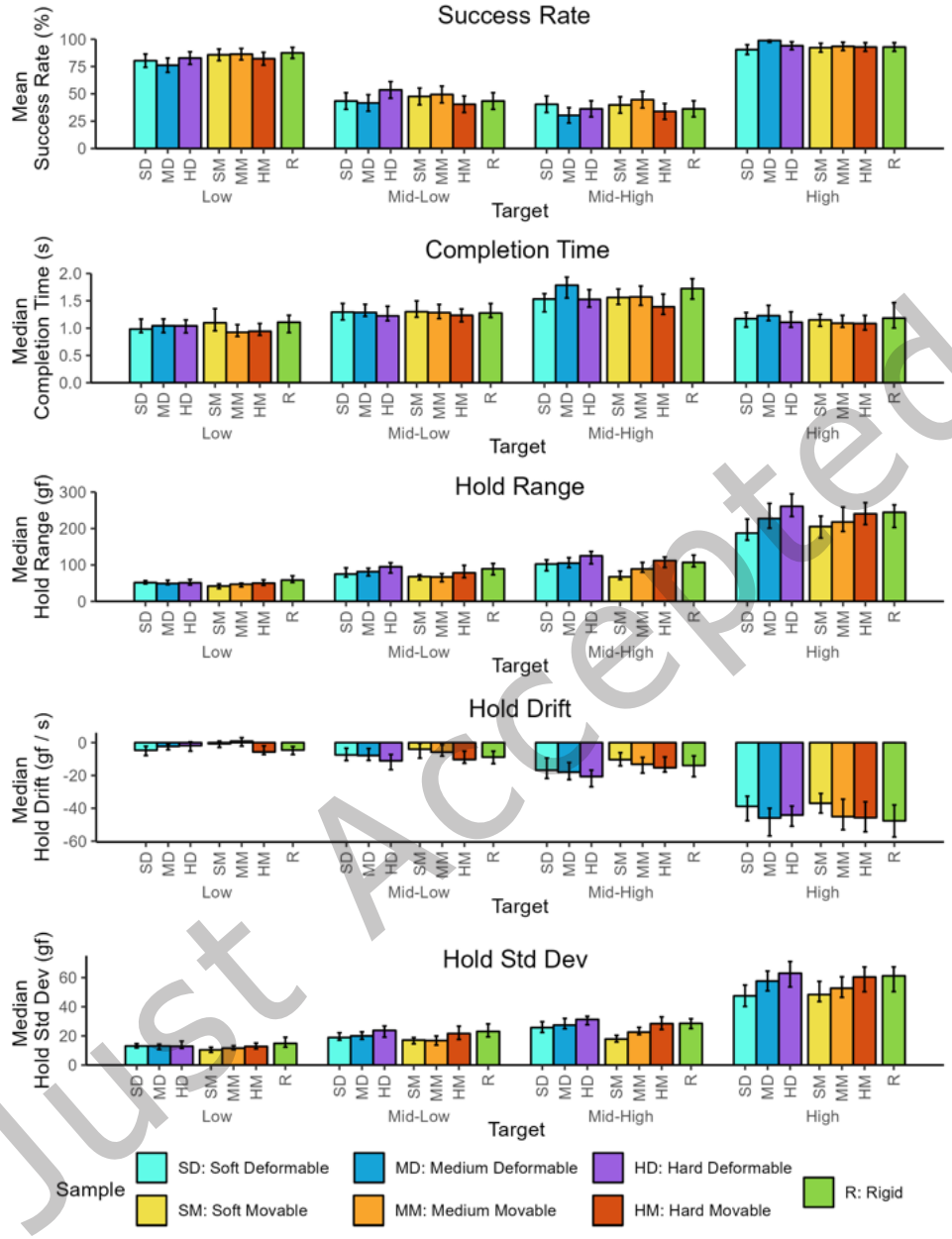

Fig. 14. Median values for (top to bottom) success rate, completion time, hold range, hold drift and hold standard deviation across each combination of SAMPLE and TARGET in Study 3. The magnitude of hold range, drift and standard deviation all consistently increase with TARGET and with STIFFNESS. The graph includes 95% upper and lower median confidence intervals.

Table 4. Significant results of the RM-ANOVA, for Study 3. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2 / F	DFn	DFd	p	R^2_{sp} / η^2
1	Success	Sample	14.61	6	162	0.024*	0.0049
2	Success	Target	1325.60	3	81	< .001***	0.2858
3	Success	Compliance Type × Stiffness	12.59	2	54	0.0018**	0.0049
4	Completion Time	Target	53.77	2.38	64.23	< .001***	0.67
5	Completion Time	Stiffness × Finger Visibility	3.30	2.00	54.00	0.044 *	0.11
6	Hold Range	Sample	8.77	6.0	162.0	< .001***	0.25
7	Hold Range	Target	288.62	2.3	61.7	< .001***	0.91
8	Hold Range	Sample × Target	1.98	18.0	486.0	0.01*	0.07
9	Hold Range	Compliance Type	12.60	1.00	27.00	0.001**	0.32
10	Hold Range	Stiffness	9.88	2.00	54.00	0.001**	0.27
11	Hold Range	Compliance Type × Target	2.86	3.00	81.00	0.042*	0.10
12	Hold Drift	Sample	2.32	6.00	162.00	0.035*	0.08
13	Hold Drift	Target	94.91	1.42	38.42	0.01*	0.78
14	Hold Drift	Compliance Type	8.32	1.00	27.00	0.008**	0.24
15	Hold SD	Sample	9.47	6.00	162.00	< .001***	0.26
16	Hold SD	Target	255.07	2.32	62.51	< .001***	0.90
17	Hold SD	Sample × Target	1.72	18.00	486.00	0.034*	0.06
18	Hold SD	Compliance Type	9.97	1.00	27.00	0.004**	0.27
19	Hold SD	Stiffness	11.80	2.00	54.00	0.001**	0.30
20	Hold SD	Compliance Type × Target	2.96	3.00	81.00	0.037*	0.10

both fixed and random effects and ($Marginal.R^2 = .347$) for just the fixed effects. The analysis showed a significant ($\chi^2_3 = 1325.6, p < .001, R^2_{sp} = .286$) main effect of TARGET on success. The post hoc tests showed a significant difference ($p < .001$) between all target levels, with performance higher on the lowest ($\mu = 84\%[81\%, 87\%]$) and highest ($\mu = 94\%[92\%, 95\%]$) targets, while being considerably lower on medium-low ($\mu = 45\%[40\%, 50\%]$) and medium-high ($\mu = 37\%[32\%, 42\%]$). The analysis also showed a significant ($\chi^2_6 = 14.611, p = .024, R^2_{sp} = .0049$) main effect of SAMPLE on success. The post hoc tests showed only a significant difference ($p = 0.040$) between medium movable ($\mu = 68\%[65\%, 72\%]$) and medium deformable ($\mu = 62\%[58\%, 65\%]$). A four-way model was also created with fixed effects of STIFFNESS, COMPLIANCE TYPE and TARGET, including an interaction between STIFFNESS, COMPLIANCE TYPE and with FINGER VISIBILITY excluded and participant number as a random effect, its explanatory power was ($Cond.R^2 = .379$) for both fixed and random effects and ($Marginal.R^2 = .345$) for just the fixed effects. The four-way model did show a significant ($\chi^2_2 = 12.591, p = .0018, R^2_{sp} = .0049$) two-way interaction between COMPLIANCE TYPE and STIFFNESS on success. There was a simple main effect ($\chi^2_1 = 6.73, p = .0095, R^2_{sp} = .0050$) of COMPLIANCE TYPE on the medium stiffness samples, with the post hoc test showing success higher on the movable ($\mu = 68\%[65\%, 72\%]$) compared with the deformable ($\mu = 62\%[58\%, 65\%]$), at this level.

6.2.2 Completion Time. Across all the trials, the median completion time was 1.25 s [1.22 s, 1.27 s]. The median completion time for each combination of SAMPLE and TARGET are shown in Figure 14, *completion time*. The three-way RM-ANOVA showed the effect of TARGET on completion time was significant ($F_{2.38,64.23} = 53.77, p < .001, \eta^2_p = .67$), with a large effect size. The post hoc tests showed significant differences ($p < 0.01$) between all levels, with completion time fastest on low ($\bar{X} = 1.02$ s [.97 s, 1.08 s] and slowest on mid-high ($\bar{X} = 1.56$ s [1.52 s,

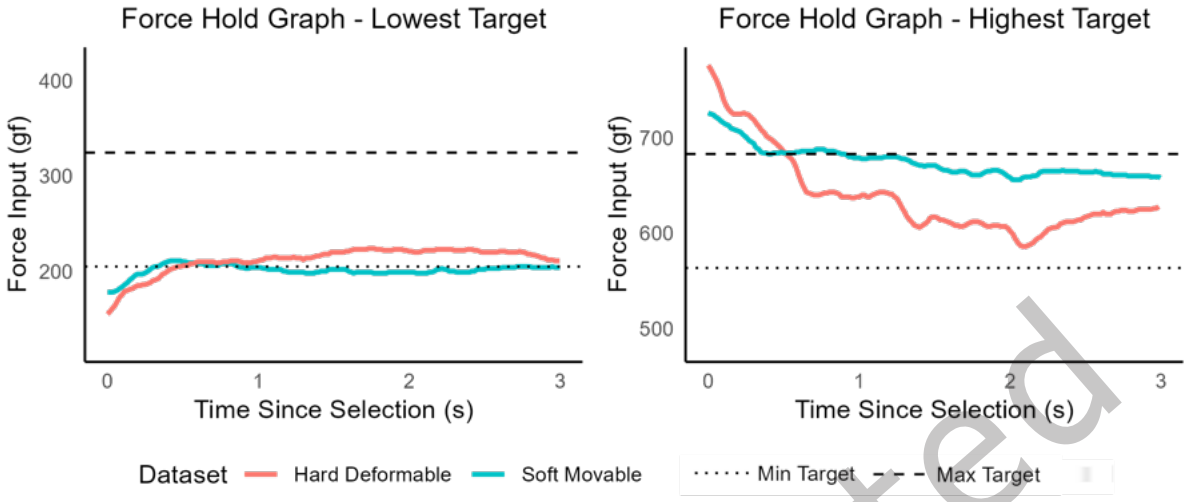

Fig. 15. Example graphs showing the applied force during the three-second hold period, for Soft Movable and Hard Deformable on (from left to right) the lowest target, and the highest target. The graphs provide an example of the lower range, drift and standard deviation at the lower force target, and the improved hold performance on softer samples.

1.63 s]). The four-way ANOVA showed a significant ($F_{2,54} = 3.3, p = .044, \eta_p^2 = .11$) two-way effect between STIFFNESS and FINGER VISIBILITY. The only significant simple main effect ($F_{1,27} = 6.72, p = .015, \eta_p^2 = .20$) was of FINGER VISIBILITY on the soft samples, with a large effect size. The post hoc tests showed completion time faster on the soft sample with the finger visible ($\bar{X}$ = 1.24 s [1.18 s, 1.30 s]) compared to without ($\bar{X}$ = 1.29 s [1.22 s, 1.39 s]).

6.2.3 Hold Range. Overall, the median hold range was 97.5 gf [94.0 gf, 101.0 gf]. The hold range results for each combination of SAMPLE and TARGET are shown in Figure 14, *hold range*. Example force-input graphs for the three-second hold period are shown in Figure 15, for different level targets and samples. The analysis showed a significant two-way interaction between TARGET and SAMPLE on hold range ($F_{18,486} = 1.98, p = .01, \eta_p^2 = .07$), with a medium effect size. There was a simple main effect of SAMPLE at all levels of the target, with hold range increasing with TARGET. For all targets combined, the hold range was smallest on soft movable ($\bar{X}$ = 76.0 gf [71.0 gf, 82.0 gf]) and highest on hard deformable ($\bar{X}$ = 116.0 gf [106.0 gf, 127.0 gf]). There was also a significant two-way interaction between TARGET and COMPLIANCE TYPE on hold range ($F_{3,81} = 2.86, p = 0.042, \eta_p^2 = .10$), with a medium effect size. There were simple main effects of COMPLIANCE TYPE for all levels of the target except high. The post hoc tests showed that for all targets combined, the range was higher on deformable ($\bar{X}$ = 103.0 gf [98.0 gf, 109.0 gf]) than movable ($\bar{X}$ = 88.0 gf [83.0 gf, 94.0 gf]). Finally, there was a significant main effect ($F_{2,54} = 9.88, p < 0.001, \eta_p^2 = .27$) of STIFFNESS on hold range, with a large effect size. Post hoc tests only showed a significant difference ($p < .001$) between the soft ($\bar{X}$ = 84.5 gf [80.0 gf, 93.0 gf]) and hard ($\bar{X}$ = 111.5 gf [104.0 gf, 118.0 gf]) samples.

6.2.4 Hold Drift. The median hold drift overall was −12.5 gf/s [−13.5 gf/s, −11.6 gf/s], and is shown for each combination of SAMPLE and TARGET in Figure 14, *hold drift*. There was a significant main effect ($F_{1.42,38.42} = 94.91, p = 0.01, \eta_p^2 = .78$) of TARGET on hold drift, with a large effect size. Post hoc tests showed significant differences ($p < .001$) between all levels, with drift smallest on the low target ($\bar{X}$ = −2.7 gf/s [−3.7 gf/s, −1.9 gf/s])

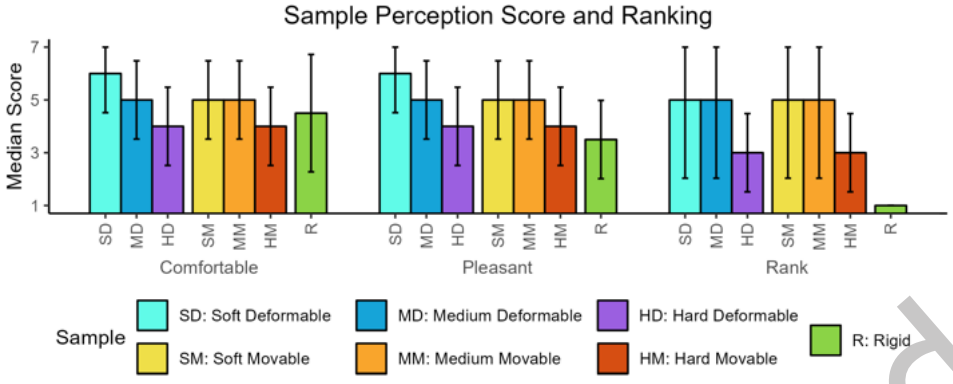

Fig. 16. Median scores for participant perceptions of how comfortable and pleasant each SAMPLE was, along with their ranking of each individual SAMPLE, for Study 3. The graph includes Mean Absolute Deviation (MAD) from the median score. Scores were given between 1 (lowest) and 7 (highest).

and largest on the high target ($\bar{X} = -43.5$ gf/s [−46.0 gf/s, −40.1 gf/s]). Along with this, there was also a significant main effect ($F_{6,162} = 2.32, p = 0.035, \eta_p^2 = .8$) of SAMPLE on hold drift, with a medium effect size. However, there were no significant differences between the samples detected from the post hoc tests. Finally, the four-way ANOVA showed a significant main effect ($F_{1,27} = 8.32, p = 0.008, \eta_p^2 = .24$) of COMPLIANCE TYPE on hold drift, with a large effect size. The post hoc tests showed the drift was larger on deformable samples ($\bar{X} = -13.6$ gf/s [−15.8 gf/s, −11.9 gf/s]) compared with movable samples ($\bar{X} = -11.4$ gf/s [−12.7 gf/s, −10.1 gf/s]) at $p = .008$.

6.2.5 Hold Standard Deviation. The median hold standard deviation was 25.1 gf/s [24.3 gf/s, 26.0 gf/s]. The hold standard deviation results for each combination of SAMPLE and TARGET are shown in Figure 14, *hold standard deviation*. There was a significant two-way interaction between TARGET and SAMPLE on hold standard deviation ($F_{18,486} = 1.72, p = 0.034, \eta_p^2 = .06$), with a medium effect size. There was a simple main effect of SAMPLE at all levels of the target, with the standard deviation increasing with target. For all targets combined, hold standard deviation was lowest on soft movable ($\bar{X} = 19.2$ gf [17.7 gf, 21.80 gf]) and highest on hard deformable ($\bar{X} = 30.0$ gf [27.7 gf, 31.9.0 gf]). The four-way RM-ANOVA revealed a significant two-way interaction between TARGET and COMPLIANCE TYPE on hold standard deviation ($F_{3,81} = 2.96, p = 0.037, \eta_p^2 = .10$), with a medium effect size. There were simple main effects of COMPLIANCE TYPE on the low and mid-high targets, with standard deviation higher on deformable for both. Finally, there was a significant main effect ($F_{2,54} = 11.8, p < 0.001, \eta_p^2 = .30$) of STIFFNESS on hold standard deviation, with a large effect size. Post hoc tests showed a significant difference ($p < .001$) between the soft ($\bar{X} = 21.8$ gf [20.3 gf, 23.5 gf]) and hard ($\bar{X} = 28.6$ gf [27.1 gf, 30.6 gf]) samples, as well as medium ($\bar{X} = 23.4$ gf [22.1 gf, 25.0 gf]) and hard.

6.2.6 Demand and Workload Survey. The NASA-TLX questions were used to judge the demand and workload of force selections on different samples and with and without their hand being obscured (see Table 13). The results of the analysis, of the demand and workload survey are shown in Table 11 and Table 12. The only significant effect seen in the Friedman tests was of SAMPLE on frustration ($\chi_6^2 = 18.969, p = .0042$). However, the post hoc tests did not show any significant differences between levels. The Friedman test revealed no significant differences for COMPLIANCE TYPE and STIFFNESS for any of the demand or workload measures.

6.2.7 Sample Perception Survey. The results of the sample perception survey for how pleasant and comfortable participants judged samples to be are shown in Figure 16, while the results of the analysis are provided in Table 14 and Table 15. Participants answered on a Likert scale from 1 (lowest) to 7 (highest).

The analysis showed a significant effect of SAMPLE ($p \leq 0.001$), STIFFNESS ($p \leq 0.005$) and COMPLIANCE TYPE ($p \leq 0.022$) on both measures. The deformable samples ($\bar{X} = 5.5$ [MAD = 2.22]) were perceived as significantly more pleasant compared with the movable samples ($\bar{X} = 5$ [MAD = 1.48]). For comfort, the deformable samples ($\bar{X} = 5$ [MAD = 1.48]) were also perceived as significantly more comfortable compared with the movable samples ($\bar{X} = 4.5$ [MAD = 2.22]). While for STIFFNESS, the soft samples ($\bar{X} = 6$ [MAD = 1.48]) were rated significantly ($p \leq 0.004$) more pleasant than both the medium ($\bar{X} = 5$ [MAD = 1.48]) and hard samples ($\bar{X} = 4$ [MAD = 1.48]). Similarly, for comfort the soft samples ($\bar{X} = 5.5$ [MAD = 2.22]) were also ranked significantly ($p \leq 0.032$) higher than both the medium samples ($\bar{X} = 5$ [MAD = 1.48]) and the hard samples ($\bar{X} = 4$ [MAD = 1.48]).

Participants' perception of sample stiffness and deformability, showed significant differences in line with expectations (i.e. soft samples were perceived to be softer than hard samples). The significant effects are reported in Table 14 and Table 15, with the results provided in Table 16.

6.2.8 Individual Sample Ranking. During the post-study interview, participants ranked the seven samples, from one (favourite) to seven (least favourite) in terms of their preference for using them during the study, with the results shown in Figure 16, while the results of the Friedman test being provided in Table 14 and Table 15. There was a significant effect of SAMPLE and STIFFNESS on ranking. As shown in the table, the rigid sample was ranked as the worst sample ($\bar{X} = 2$, MAD = 0.0), while softer samples were ranked higher and harder samples lower, demonstrating a clear participant preference.

6.2.9 Post-Study Interview. The approach to this thematic analysis mirrors Study 2, breaking it into the following three themes: **Perceived Performance**, **Perceived Feeling**, and **Sample Perception**. In total, 100 quotes were extracted from the semi-structured interview transcripts.

Perceived performance. Nearly half of the participants (n = 12) described how a sample affected task difficulty. Eight participants noted a sample made the task easier, with soft (3) being mentioned the most. In contrast, six participants commented that the task was more challenging due to the sample, with hard (2) and rigid samples (3) mentioned most often.

Over half of the participants (n = 16) explained how the sample influenced their ability to perceive their applied force, using terms such as *"force awareness"* or *"feedback"*. Soft was seen to benefit force awareness the most (4), while rigid was seen to negatively impact force awareness (6). The most notable results were for surface type, with movable consistently described as improving force awareness (7), for instance: *"I could kind of gauge what amount of pressure I was adding on, because I could just feel it move"* (P27). The impact of the sample on *"accuracy"* was mentioned by five participants, with four of them ascribing improved accuracy due to the movable samples.

Overall, participants perceived their performance as enhanced on softer samples and reduced on harder and rigid samples. In particular, they perceived soft and movable samples as providing more feedback and force awareness. This aligns with the responses from the individual sample ranking in subsubsection 6.2.8, in which participants consistently ranked soft samples highly. However, these results were not observed in the quantitative data for the selection phase.

Perceived Feeling. Participants also ranked samples based on how they felt. Nearly half of them described a sample as *"comfortable"* (11), while only five described a sample as *"uncomfortable"*. The soft sample was most frequently perceived as comfortable (6), followed by the deformable surface (4) and the medium (4). While rigid (2) and hard (4) were most commonly described as uncomfortable. Eight participants made general positive comments about the samples, describing them as *"nice"* or *"liking"* them. The soft and deformable samples were

each mentioned by five of these participants, making comments such as *"I like the soft one, but I don't know if that was in terms of how useful it was. It was more just fun to push."* (P1) and *"I think it was just sort of, how nice it was for my finger to move through"* (P26). Finally, only two participants described the samples as *"fatiguing"* or *"painful"*, though these comments were exclusively directed at the hard (2) and rigid (1) samples. Overall, participants show consistent preferences for softer and deformable samples, describing them as *"comfortable"* and *"nice"*, in contrast to the hard and rigid samples which are viewed negatively and described as *"painful"*, *"fatiguing"*, and *"uncomfortable"*.

Sample Perception. Five participants described the medium sample as having a good balance, making comments similar to (P22) *"I prefer the medium because it's not too soft, not too hard."* or (P28) *"I suppose it's a mixture of the comfort and feeling of control."* Seven participants referred to the travel of a sample. Three participants discussed the travel of the soft samples positively, such as *"There's more freedom to move, more feedback"* (P5). In contrast, two participants described disliking the amount of travel on the soft samples, including *"The soft I think it gave way a bit too much"* (P11). Three mentioned movable positively, for instance, *"the entire thing goes down rather than just, a tiny bit"* (P18).

Summary. This analysis highlights the preferences expressed by participants within the sample ranking, in which soft samples were ranked top and the hard and rigid samples were worse. While participants generally felt their performance improved with soft and movable samples, there were notable variations between participants. Nonetheless, when it comes to the feeling of samples, soft samples were clearly preferred over hard samples, which strongly aligns with the overall ranking of the samples.

6.3 Summary and Discussion

Study 3 was designed and carried out to address **RQ3** which was broken into **RQ3.1–3.3**. *RQ3.1* considered the impact of the SAMPLE, TARGET and FINGER VISIBILITY on force-selection performance (success and completion time) without visual feedback. The results showed a significant effect of SAMPLE on success, however this was limited to higher success rates for movable compared with deformable on the medium stiffness samples. The only effect of STIFFNESS on completion time, was that it was lower for softer samples during the finger visible condition. The impact of TARGET on performance was significant, with both higher success rates and lower completion times on the outer two unbounded cells (high and low), compared with the inner bounded cells. Performance on the mid-high cell was worse across both measures, while success peaked on the high target and completion time was faster on the low target. The effect of FINGER VISIBILITY was limited, with the only significant effect being faster completion times on the soft samples when the participant could see their finger. *RQ3.2* addressed the user's ability to maintain a constant force input after selection, without visual feedback. The range and standard deviation of force values when holding followed similar patterns, increasing with STIFFNESS and TARGET. Along with this range and standard deviation were higher on the deformable samples for all the targets apart from high. The drift of the hold values was impacted by the sample COMPLIANCE TYPE, and was larger on deformable surfaces. Drift also increased with TARGET, ranging from nearly zero ($\bar{X} = -2.7$ gf/s [−3.7 gf/s, −1.9 gf/s]) at the low target to −43.5 gf/s [−46.0 gf/s, −40.1 gf/s] on the high target. FINGER VISIBILITY had no significant effect on input performance while maintaining force. Combined with the selection results, this suggests that performance remains consistent regardless of the user's perception of their finger. Finally, for *RQ3.3*, our results show that SAMPLE has a limited impact on workload and demand, however there was a significant effect of SAMPLE, STIFFNESS and COMPLIANCE TYPE on perceived comfort and pleasantness. Participants found deformable and softer samples to be more comfortable and pleasant and tended to view force-awareness and feedback as improved on soft samples but worsened on hard samples. Within the study, the rigid and hard samples were ranked lowest while the softer and deformable samples were ranked higher. Participants also described the physical properties

of the softer and deformable samples more positively, while the rigid and hard samples viewed more negatively, using words such as *"fatiguing"* and *"uncomfortable"*.

6.3.1 Impact of Sample on Selection and Hold. The study revealed that SAMPLE affected success, with post hoc tests showing a significantly higher success rate for the medium stiffness movable sample compared with the equivalent deformable. One potential explanation for this is that the increased kinesthetic and cutaneous feedback from the deformable sample may be distracting or overwhelming, making it harder to recall the different levels. However, further investigation is needed to confirm this and consider why this trend is not seen at other stiffness levels.

The study also showed two examples of how sample significantly impacts hold performance, both of which are significant and previously unreported. First, softer samples resulted in lower range and standard deviation of values during the hold period, and second, all measures of force control during the hold phase improved on movable surfaces, including the total drift during the hold period. We speculate that the effect of STIFFNESS on hold range and standard deviation could be caused by damping out micro-vibrations in a user's finger, resulting in less transfer of force input to the device. This would explain why STIFFNESS has an effect on range and standard deviation, but not drift. The reason for the effect of COMPLIANCE TYPE on all three hold measures is less clear. It could be that the one-dimensional mechanical movement of the movable device provides more focused feedback compared with the more complex feedback for deformable which is in multiple dimensions and depends on the direction of the force vector. This may be exacerbated by small changes when pressing into the surface of the deformable samples that are distracting or inconsistent between inputs.

6.3.2 User Preferences for Soft and Deformable Samples. Study 3 demonstrated a clear user preference for soft, deformable samples during force-input interactions without visual feedback, echoing findings from Study 2. These samples reduce discomfort and fatigue without affecting selection performance, although participants perceived better force awareness on soft and movable samples, despite actual selection performance not aligning with these perceptions. This relationship between stiffness and force awareness is logical, as softer samples enhance kinesthetic feedback. However, the increased force awareness on movable samples compared with deformable is less expected and suggests that physical movement may be easier to perceive than deformation. The divergence between actual performance and perceptions of performance should be considered carefully. Although participants perceived force awareness as best on soft and movable samples, we did not see this trend in the success rate of the task. However, improving participants' perception of their ability is a crucial factor in the design of input devices. This is further evidence indicating that soft samples should be preferred when designing force-input devices, while the surface type of the devices should be selected based on the balance between comfort and pleasantness versus perceptions of performance and awareness.

6.3.3 Limitations of Eyes-free Force Input. The limited performance of force input interactions without visual feedback is a significant limiting factor in their wider adoption. This study aligns with prior work [60, 91], which showed that practical selection performance is restricted to two to three levels, depending on accuracy requirements. It has shown that within this context, conducting force input on compliant surfaces is not able to overcome this limitation. The ability of a user to make multi-level discrete gaze-free selections would greatly expand their application space. Achieving this may still be possible, but by using alternative approaches. It could be possible to use haptic feedback, such as signifying levels with vibrotactile or force-feedback clicks [88, 105], or audio-cues. Alternatively, relative force-input interactions, where users make relative changes from an arbitrary starting point to reach different levels, instead of absolute force values relating to specific input levels, could be used. This would leverage a user's ability to maintain a force value without visual feedback, especially on soft and movable surfaces, to make accurate selections.

Recommendations. We present **three recommendations** for force input interactions without visual or any other form of feedback, based on our findings from Study 3:

(1) *Force hold interactions without feedback are best suited to soft and movable surfaces, which reduce the total range, drift and standard deviation during the force input.*
(2) *Where possible, force input interactions without feedback should be conducted on softer surfaces to improve the experience. Deformable surfaces can enhance comfort and pleasantness, and movable surfaces increase the perceived performance.*
(3) *To ensure accurate and consistent force inputs without feedback, the number of input levels should be limited to three.*

6.3.4 Limitations. Our selection of the target cell boundaries is a limitation of this study; while un-bounding the top and bottom cells increases the ecological validity of the design, it does have drawbacks. First, as the selection performance was greatly improved on the unbounded cells, this made it very difficult to detect the impact of target force on performance. Based on the results of Study 2, we would have expected performance to be worse at the highest force input level, instead, performance was best at this point. The second challenge is that participants may have been applying significantly different forces, due to the lack of visual feedback, something which would have been more prevalent on the top and bottom cells. This has a particular impact on the hold interactions because the applied force at the beginning of the hold may vary between participants, this could impact hold performance. A future study, similar to Chen [12], in which visual feedback is provided to allow the participant to apply the correct starting force before it's removed during the hold phase could address this limitation. Finally, although there was no feedback of any type during Study 3, the lack of visual feedback has been focused on, because of the contrast with Study 2. It may be possible to include other types of feedback, such as audio or haptic to improve performance without visual feedback.

7 DISCUSSION, LIMITATIONS, AND FUTURE WORK

This work systematically investigated the impact of stiffness and compliance type on force-input interactions. Within this section, we consolidate the findings from the three studies and discuss their impact on the design of force-input interactions and compliant devices. This section is organised into three segments: i) Beyond the Rigid Force-Input Device Paradigm, ii) Reshaping our Understanding of Compliance Characteristics, and iii) Designing Better Force-Input Interactions.

7.1 Beyond the Rigid Force-Input Device Paradigm

Across Studies 2 and 3, the rigid surface was consistently outperformed in both user preference and force-hold performance, while soft surfaces performed best. These findings challenge the current focus on rigid devices for force-input interactions [91, 92, 97]. Rather than improving force-input interactions, our studies show how rigid surfaces actually increase discomfort and fatigue and reduce participants' ability to maintain a constant force. ***Based on our results, soft and deformable surfaces should be considered for force-input selection using a finger.*** Instead, extending force-input interactions to compliant surfaces, especially for force-hold interactions, can enhance both performance and user experience. It should be noted these results are for a single use case of a user applying force onto a flat horizontal device, further testing is required for different contexts, such as holding a device, as well as standing or walking [82]. This shift from the rigid force-input device paradigm poses several challenges and questions. First, rigid devices are more accessible and cost-effective than complex tangible compliant devices [38], creating a tradeoff between the interaction benefits of compliant devices and the ubiquity of rigid ones. However, compliant interaction devices [23, 54, 103] are becoming prevalent, with movable examples already common in our everyday lives [1, 9, 67]. Furthermore, advances in ubiquitous sensing technologies could also facilitate this shift by leveraging the compliant devices we already use daily as untapped

soft interfaces for force input, such as furniture [11] or wearables [62, 94]. Thus, despite initial challenges, the rise of soft interfaces and technological advances are making compliant devices increasingly accessible and integrated into daily life. Resulting in the move away from rigid devices and towards compliant devices becoming more feasible.

Second, although these studies focus on force input applied with the index finger, our findings could also be relevant to a variety of other force-input techniques. As demonstrated by Taher et al. [82], the finger or hand used and a user's posture can significantly impact the force input interaction. Force input is also not limited to normal pressure; shearing [31, 40], pinching [66], and bending [10] are viable force input methods. We believe that the benefits of compliant devices during normal force input, such as tactile feedback and increased finger-pad contact, may extend to these other contexts, offering similar improvements. Future studies should investigate a diverse array of force input techniques on compliant surfaces, comparing them with rigid devices, to fully uncover the potential and advantages of these underutilised interfaces.

Finally, there is a rising trend in leveraging pseudo-haptics to simulate compliance through methods such as manipulating the control-to-display ratio [78, 95], using electrotactile grains [37, 42] or with skin-streching [55]. Additionally, there is potential to create semi-compliant devices by overlaying a thin, compliant layer on rigid surfaces, thereby introducing a degree of softness to existing rigid devices [103]. The relationship between our findings and the potential of pseudo-haptics or semi-compliant devices is promising but still needs further exploration. Future research could focus on evaluating whether these methods can achieve the same benefits we observed with fully compliant surfaces. In particular, they should investigate if similar improvements in hold force control and user preference can be achieved.

7.2 Reshaping our Understanding of Compliance Characteristics

The results from all three studies shed light on previously unreported impacts of compliance characteristics. Study 1 found that deformable surfaces allowed for higher maximum comfortable force compared to movable ones. Studies 2 and 3 revealed that soft surfaces reduced hold range and drift compared to hard surfaces, with Study 3 demonstrating that movable surfaces outperformed deformable ones on all hold measurements. Soft and deformable surfaces also performed best in surveys and post-study interviews. As a result of these distinct differences, it is important that we reshape our understanding of compliant devices.

7.2.1 Movable vs Deformable: Examining the Differences. Previous work has often treated movable and deformable devices interchangeably [6, 56, 61, 84]. However, our findings reveal significant differences between them, showing how movable and deformable devices provide distinct user experiences and impact performance differently. Researchers should avoid equating or grouping them without careful consideration. While Boem and Troiano [6] highlighted the need to differentiate in the context of "deformable inputs", our work underscores the larger differences between movable and deformable devices. Deformable devices can expand the comfortable force input range and enhance user experience, whereas movable devices excel at providing precise force-hold inputs. Therefore, we propose that ***it is essential that designers focus on desired interaction characteristics when choosing the type of compliance for a force-input device.***

Although these studies have demonstrated the differences between movable and deformable in one context, the choice between them is more complex. First, movable devices have distinct benefits which have not been explored in these studies, such as their improved capability of rendering detailed haptic feedback [28, 88, 104]. Furthermore, movable devices are more common and easier to design and build, as they align with component-centric design [29]. However, recent advancements in material-centric approaches [30, 75, 96] and 3D printing [64] technologies are making soft, elastic interactive devices more accessible. As these trends continue, deformable devices may become more common in our everyday lives, and we can further leverage their unique advantages.

7.2.2 Leveraging the Benefits of Variable-Stiffness Devices. Variable stiffness devices can dynamically change their stiffness, which is normally used to provide haptic feedback [18, 19] or to simulate objects [69], materials [47, 51] or even medical procedures [73, 85]. Variable stiffness can also create passive haptics, which has been used in improving input performance in devices such as sliders [45], dials and knobs [5, 86]. Notably, Studies 2 and 3 demonstrate that varying stiffness can significantly enhance force input interactions, improving performance and user experience, such as, with soft samples offering greater comfort, pleasantness, and better hold control. This opens up a novel way of utilising these types of devices, changing stiffness to facilitate specific interactions. We believe this shows how ***variable-stiffness devices can be leveraged to maximise user performance and experience during force-input.*** Building on prior work, which has demonstrated how stiffness can impact the interaction potential [43, 87] and users' emotional response [74, 76], our results continue to expand the potential of variable stiffness devices. Furthermore, hard and rigid surfaces are likely better suited for traditional touch-screen gestures, such as swiping or pinching, which users are accustomed to on glass screens. Therefore, to improve user experience, consider designing flat solid surfaces that can dynamically adjust their stiffness. For example, such surfaces could become softer at specific locations or under the user's finger to support more accurate force input and reduce discomfort. After the interaction, the surface could return to its original stiffness or adjust to provide haptic feedback, indicating that the input is complete. Further investigation into dynamic compliance could reveal novel ways to leverage the advantages of variable stiffness devices for improved force input interactions.

*7.2.3 Dynamic Stiffness **during** Interaction.* With the increasing prevalence of variable-stiffness devices, it will become important to understand the impact of stiffness change during force-input interactions. Within these studies, the sample properties were only changed between interactions when a participant was not touching the device, however, some devices [50, 54] can alter the stiffness while the user is applying force. Future work should evaluate how this impacts user force input and whether haptic force feedback can be provided during a force input interaction.

7.3 Designing Better Force-Input Interactions

Throughout the three studies, the results have demonstrated the impact of variables such as visual feedback and force range on user performance. We synthesise the findings from across the three studies to aid the design of force-input interactions more broadly.

7.3.1 The Impact of Input Range on Performance. In Studies 2 and 3, increasing applied force consistently reduced performance, with the most significant reductions at the highest force levels. These findings align with Schmitz et al. [66] and Mizobuchi et al. [49], who both showed the detrimental impact of increased input force on the force-input interaction. We attribute this performance decline to the increased physical demand of the input task as the applied force rises, making it more challenging. Study 1 also revealed substantial variation in comfortable force values between participants, which is likely caused by individual differences. Together, these finding strongly suggests that ***force-input range should be user-defined and lowered where possible to avoid performance breakdown seen at higher input forces.*** Our studies did not account for individual participant factors, such as fingertip surface area, sensitivity, or finger and arm strength. Future research should investigate the impact of these factors on both the comfortable force input range and the relationship between target force and performance. This can be used to tailor force-input interactions to individual user differences and improve performance and experience further.

7.3.2 Force-Input Selection with and Without Visual Feedback. This work adds to the existing literature [58, 91], which highlights the limitation of force-input without visual feedback. With visual feedback, participants could select the target at least half the time on the 60-level scale. In contrast, when selecting from just four levels

without visual feedback, participants' success on the mid-high target was below 40%, although the performance was much higher when selecting the unbounded upper and lower cells. Combined with the prior work, this indicates that without additional feedback, force input selection without visual feedback is practically limited to two or three levels, depending upon the accuracy required. The results from Studies 2 and 3 show that the additional kinesthetic and cutaneous feedback from compliant samples is not enough to overcome this reliance on visual feedback, and force-input without visual feedback is still limited. While tactile cues can provide valuable feedback, relying solely on users to remember specific force levels is challenging. We still believe that ***achieving reliable force input without visual feedback can lead to interesting and novel opportunities for gaze-free force inputs, however additional feedback modalities are likely required to enhance the user experience***. Future research should explore alternative methods, and one promising avenue is the integration of multimodal feedback such as vibrotactile clicks or dynamic force-feedback [28].

8 CONCLUSION

Within this paper, we conducted three comparative studies investigating how compliance and stiffness affect force-input selection. This work is the first to directly compare deformable, movable, and rigid samples, and to investigate the impact of sample compliance on force input without visual feedback. We quantify lower (205 gf [196 gf, 216 gf]) and higher (439 gf [410 gf, 467 gf]) comfortable force levels and identified a statistically significant increase (17.2%) in the higher comfortable force applied to deformable samples compared with movable samples. This is important to consider when designing force-input interactions because what is comfortable on one type of compliance may not be on another. Although traditionally force-input interactions are studied and deployed on rigid surfaces, our results demonstrate that they are the least favoured surface which are consistently ranked lowest and described negatively. In addition, the stiffness of a surface can significantly improve force-input interactions when users have visual feedback, with softer samples reducing the user force input range and standard deviation compared with harder samples. Similar trends are also present with force-input without visual feedback. Softer samples reduce the range and standard deviation while holding a force input. At the same time, movable samples performed better when holding force-input interactions than deformable samples. Despite this, compliant surfaces still suffered from severe performance limitations when providing force-input without any visual feedback, with accurate selection limited to two or three items. This needs to be addressed for this to be a viable interaction technique in the future. Overall, these previously unreported differences between movable, deformable and rigid surfaces on force-input interactions should be taken into account when designing force-input interactions and how we view the relationship between force-input interactions and compliant devices.

ACKNOWLEDGMENTS

We would like to thank Cameron Steer for their invaluable discussions and insightful thoughts during the early stages of this project. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (FORCE-UI, Grant agreement No.853063).

REFERENCES

[1] Jason Alexander, John Hardy, and Stephen Wattam. 2014. Characterising the Physicality of Everyday Buttons. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (Dresden, Germany) *(ITS '14)*. Association for Computing Machinery, New York, NY, USA, 205–208. https://doi.org/10.1145/2669485.2669519

[2] Jason Alexander, Anne Roudaut, Jurgen Steimle, Kasper Hornbæk, Miguel Bruns Alonso, Sean Follmer, and Timothy Merritt. 2018. Grand Challenges in Shape-Changing Interface Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 299:1–299:14. https://doi.org/10.1145/3173574.3173873

[3] Axel Antoine, Sylvain Malacria, and Géry Casiez. 2017. ForceEdge: Controlling Autoscroll on Both Desktop and Mobile Computers Using the Force. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*.

Association for Computing Machinery, New York, NY, USA, 3281–3292. https://doi.org/10.1145/3025453.3025605
[4] Apple. 2023. How to use the Force Touch trackpad. https://support.apple.com/en-gb/102309
[5] Andrea Bianchi, Ian Oakley, Jong Keun Lee, and Dong Soo Kwon. 2010. The haptic wheel: design & evaluation of a tactile password system. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 3625–3630. https://doi.org/10.1145/1753846.1754029
[6] Alberto Boem and Giovanni Maria Troiano. 2019. Non-Rigid HCI: A Review of Deformable Interfaces and Input. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 885–906. https://doi.org/10.1145/3322276.3322347
[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806
[9] Michael Brown, Aidan Kehoe, Jurek Kirakowski, and Ian Pitt. 2015. *Beyond the Gamepad: HCI and Game Controller Design and Evaluation*. Springer International Publishing, Cham, 263–285. https://doi.org/10.1007/978-3-319-15985-0_12
[10] Jesse Burstyn, Juan Pablo Carrascal, and Roel Vertegaal. 2016. Fitts' Law and the Effects of Input Mapping and Stiffness on Flexible Display Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3649–3658.
[11] Michael Chamunorwa, Mikołaj P. Wozniak, Sarah Vöge, Heiko Müller, and Susanne C.J. Boll. 2022. Interacting with Rigid and Soft Surfaces for Smart-Home Control. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 211 (sep 2022), 22 pages. https://doi.org/10.1145/3546746
[12] Jyh-shing Chen. 1998. *Human haptic interaction with soft objects: discriminability, force control, and contact visualization*. Ph. D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA, USA. Advisor(s) Srinivasan, Mandayam A. OCLC 37537724.
[13] Jacob Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Routledge, New York.
[14] Sony Corporation. 2020. *DualSense Wireless Controller: The Innovative New Controller for PS5*. Sony Corporation. https://www.playstation.com/en-gb/accessories/dualsense-wireless-controller/
[15] Christian Corsten, Marcel Lahaye, Jan Borchers, and Simon Voelker. 2019. ForceRay: Extending Thumb Reach via Force Input Stabilizes Device Grip for Mobile Touch Input. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300442
[16] Christian Corsten, Simon Voelker, and Jan Borchers. 2017. Release, Don't Wait! Reliable Force Input Confirmation with Quick Release. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (Brighton, United Kingdom) *(ISS '17)*. Association for Computing Machinery, New York, NY, USA, 246–251. https://doi.org/10.1145/3132272.3134116
[17] Christian Corsten, Simon Voelker, Andreas Link, and Jan Borchers. 2018. Use the Force Picker, Luke: Space-Efficient Value Input on Force-Sensitive Mobile Touchscreens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3174235
[18] Antoine Costes, Fabien Danieau, Ferran Argelaguet-Sanz, Anatole Lécuyer, and Philippe Guillotel. 2018. KinesTouch: 3D Force-Feedback Rendering for Tactile Surfaces. In *Virtual Reality and Augmented Reality*, Patrick Bourdot, Sue Cobb, Victoria Interrante, Hirokazu kato, and Didier Stricker (Eds.). Springer International Publishing, Cham, 97–116.
[19] Alexandra Delazio, Ken Nakagaki, Roberta L. Klatzky, Scott E. Hudson, Jill Fain Lehman, and Alanson P. Sample. 2018. Force Jacket: Pneumatically-Actuated Jacket for Embodied Haptic Experiences. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173894
[20] W. J. Dixon. 1960. Simplified Estimation from Censored Normal Samples. *The Annals of Mathematical Statistics* 31, 2 (1960), 385 – 391. https://doi.org/10.1214/aoms/1177705900
[21] Pierre Dragicevic. 2016. *Fair Statistical Communication in HCI*. Springer International Publishing, Cham, 291–330. https://doi.org/10.1007/978-3-319-26633-6_13
[22] Zhuzhi Fan and Céline Coutrix. 2023. Impact of Softness on Users' Perception of Curvature for Future Soft Curvature-Changing UIs. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 747, 19 pages. https://doi.org/10.1145/3544548.3581179
[23] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: Dynamic Physical Affordances and Constraints Through Shape and Object Actuation. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 417–426. https://doi.org/10.1145/2501988.2502032
[24] Bruno Fruchard, Paul Strohmeier, Roland Bennewitz, and Jürgen Steimle. 2021. Squish This: Force Input on Soft Surfaces for Visual Targeting Tasks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 219, 9 pages. https://doi.org/10.1145/3411764.3445623

[25] Dhirendra Ghosh and Andrew Vogt. 2012. Outliers: An Evaluation of Methodologies. In *Proceedings of the Joint Statistical Meetings*. American Statistical Association, San Diego, CA, 3455–3460. Section on Survey Research Methods.

[26] Ryota Gomi, Ryo Suzuki, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. 2024. InflatableBots: Inflatable Shape-Changing Mobile Robots for Large-Scale Encountered-Type Haptics in VR. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 423, 14 pages. https://doi.org/10.1145/3613904.3642069

[27] Jaehyun Han, Seongkook Heo, Jiseong Gu, and Geehyuk Lee. 2014. Trampoline: A Double-Sided Elastic Touch Device for Repoussé and Chasing Techniques. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 1627–1632. https://doi.org/10.1145/2559206.2581252

[28] Teng Han, Qian Han, Michelle Annett, Fraser Anderson, Da-Yuan Huang, and Xing-Dong Yang. 2017. Frictio: Passive Kinesthetic Force Feedback for Smart Ring Output. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 131–142. https://doi.org/10.1145/3126594.3126622

[29] Ollie Hanton, Mike Fraser, and Anne Roudaut. 2024. DisplayFab: The State of the Art and a Roadmap in the Personal Fabrication of Free-Form Displays Using Active Materials and Additive Manufacturing.. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 333, 24 pages. https://doi.org/10.1145/3613904.3642708

[30] Ollie Hanton, Michael Wessely, Stefanie Mueller, Mike Fraser, and Anne Roudaut. 2020. ProtoSpray: Combining 3D Printing and Spraying to Create Interactive Displays with Arbitrary Shapes. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376543

[31] Chris Harrison and Scott Hudson. 2012. Using Shear as a Supplemental Two-Dimensional Input Channel for Rich Touchscreen Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Austin Texas USA, 3149–3152. https://doi.org/10.1145/2207676.2208730

[32] Chris Harrison and Scott E. Hudson. 2009. Providing Dynamically Changeable Physical Buttons on a Visual Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 299–308. https://doi.org/10.1145/1518701.1518749

[33] Liang He, Gierad Laput, Eric Brockmeyer, and Jon E. Froehlich. 2017. SqueezaPulse: Adding Interactive Input to Fabricated Objects Using Corrugated Tubes and Air Pulses. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17)*. ACM, New York, NY, USA, 341–350. https://doi.org/10.1145/3024969.3024976

[34] Seongkook Heo, Jaehyun Han, and Geehyuk Lee. 2015. Designing for Hover- and Force-Enriched Touch Interaction. In *Computer-Human Interaction. Cognitive Effects of Spatial Interaction, Learning, and Ability: 25th Australian Computer-Human Interaction Conference, OzCHI 2013, Adelaide, SA, Australia, November 25-29, 2013, Revised and Extended Papers* (Adelaide, SA, Australia). Springer-Verlag, Berlin, Heidelberg, 68–87. https://doi.org/10.1007/978-3-319-16940-8_4

[35] Seongkook Heo and Geehyuk Lee. 2011. Force Gestures: Augmenting Touch Screen Gestures with Normal and Tangential Forces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11*. ACM Press, Santa Barbara, California, USA, 621. https://doi.org/10.1145/2047196.2047278

[36] Seongkook Heo and Geehyuk Lee. 2012. ForceDrag: Using Pressure as a Touch Input Modifier. In *Proceedings of the 24th Australian Computer-Human Interaction Conference on - OzCHI '12*. ACM Press, Melbourne, Australia, 204–207. https://doi.org/10.1145/2414536.2414572

[37] Seongkook Heo, Jaeyeon Lee, and Daniel Wigdor. 2019. PseudoBend: Producing Haptic Illusions of Stretching, Bending, and Twisting Using Grain Vibrations. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 803–813. https://doi.org/10.1145/3332165.3347941

[38] Lars Erik Holmquist. 2023. Bits are Cheap, Atoms are Expensive: Critiquing the Turn Towards Tangibility in HCI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 409, 8 pages. https://doi.org/10.1145/3544549.3582744

[39] Torsten Hothorn, Kurt Hornik, Mark A Van De Wiel, and Achim Zeileis. 2008. Implementing a class of permutation tests: the coin package. *Journal of statistical software* 28, 8 (2008), 1–23.

[40] Mengting Huang, Kazuyuki Fujita, Kazuki Takashima, Taichi Tsuchida, Hiroyuki Manabe, and Yoshifumi Kitamura. 2019. ShearSheet: Low-Cost Shear Force Input with Elastic Feedback for Augmenting Touch Interaction. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 77–87. https://doi.org/10.1145/3343055.3359717

[41] Yvonne Jansen, Thorsten Karrer, and Jan Borchers. 2010. MudPad: A Tactile Memory Game. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 306–306. https://doi.org/10.1145/1936652.1936734

[42] Arata Jingu, Nihar Sabnis, Paul Strohmeier, and Jürgen Steimle. 2024. Shaping Compliance: Inducing Haptic Illusion of Compliance in Different Shapes with Electrotactile Grains. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 432, 13 pages. https://doi.org/10.1145/3613904.3641907

[43] Sang-Su Lee, Sohyun Kim, Bopil Jin, Eunji Choi, Boa Kim, Xu Jia, Daeeop Kim, and Kun-pyo Lee. 2010. How users manipulate deformable displays as input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1647–1656. https://doi.org/10.1145/1753326.1753572

[44] Hongnan Lin, Liang He, Fangli Song, Yifan Li, Tingyu Cheng, Clement Zheng, Wei Wang, and HyunJoo Oh. 2022. FlexHaptics: A Design Method for Passive Haptic Inputs Using Planar Compliant Structures. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 169, 13 pages. https://doi.org/10.1145/3491102.3502113

[45] Lars Lischke, Paweł W. Woźniak, Sven Mayer, Andreas Preikschat, and Morten Fjeld. 2017. Using Variable Movement Resistance Sliders for Remote Discrete Input. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (Brighton, United Kingdom) *(ISS '17)*. Association for Computing Machinery, New York, NY, USA, 116–125. https://doi.org/10.1145/3132272.3134135

[46] Ross McLachlan, Daniel Boland, and Stephen Brewster. 2014. Transient and Transitional States: Pressure as an Auxiliary Input Modality for Bimanual Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 401–410. https://doi.org/10.1145/2556288.2557260

[47] Mustafa Mete, Haewon Jeong, Wei Dawid Wang, and Jamie Paik. 2024. SORI: A softness-rendering interface to unravel the nature of softness perception. *Proceedings of the National Academy of Sciences* 121, 13 (2024), e2314901121.

[48] Takashi Miyaki and Jun Rekimoto. 2009. GraspZoom: zooming and scrolling control model for single-handed mobile interaction. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Bonn, Germany) *(MobileHCI '09)*. Association for Computing Machinery, New York, NY, USA, Article 11, 4 pages. https://doi.org/10.1145/1613858.1613872

[49] Sachi Mizobuchi, Shinya Terasaki, Turo Keski-Jaskari, Jari Nousiainen, Matti Ryynanen, and Miika Silfverberg. 2005. Making an impression: force-controlled pen input for handheld devices. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1661–1664. https://doi.org/10.1145/1056808.1056991

[50] Ken Nakagaki, Daniel Fitzgerald, Zhiyao (John) Ma, Luke Vink, Daniel Levine, and Hiroshi Ishii. 2019. inFORCE: Bi-directional 'Force' Shape Display for Haptic Interaction. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '19)*. Association for Computing Machinery, New York, NY, USA, 615–623. https://doi.org/10.1145/3294109.3295621

[51] Ken Nakagaki, Luke Vink, Jared Counts, Daniel Windham, Daniel Leithinger, Sean Follmer, and Hiroshi Ishii. 2016. Materiable: Rendering Dynamic Material Properties in Response to Direct Physical Touch with Shape Changing Interfaces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2764–2772. https://doi.org/10.1145/2858036.2858104

[52] Shinichi Nakagawa and Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution* 4, 2 (2013), 133–142.

[53] James David Nash, Kim Sauvé, Adwait Sharma, Christopher Clarke, and Jason Alexander. 2025. Dataset for "Investigating the Impact of Deformable, Movable and Rigid Surfaces on Force-Input Interactions". https://doi.org/10.15125/BATH-01522

[54] James David Nash, Cameron Steer, Teodora Dinca, Adwait Sharma, Alvaro Favaratto Santos, Benjamin Timothy Wildgoose, Alexander Ager, Christopher Clarke, and Jason Alexander. 2024. DeformIO: Dynamic Stiffness Control on a Deformable Force-sensing Display. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 98, 8 pages. https://doi.org/10.1145/3613905.3650772

[55] Zhan Fan Quek, Samuel B. Schorr, Ilana Nisky, Allison M. Okamura, and William R. Provancher. 2014. Augmentation Of Stiffness Perception With a 1-Degree-of-Freedom Skin Stretch Device. *IEEE Transactions on Human-Machine Systems* 44, 6 (2014), 731–742. https://doi.org/10.1109/THMS.2014.2348865

[56] Raf Ramakers, Johannes Schöning, and Kris Luyten. 2014. Paddle: highly deformable mobile devices with physical controls. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2569–2578. https://doi.org/10.1145/2556288.2557340

[57] Gonzalo Ramos and Ravin Balakrishnan. 2005. Zliding: Fluid Zooming and Sliding for High Precision Parameter Manipulation. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. Association for Computing Machinery, New York, NY, USA, 143–152. https://doi.org/10.1145/1095034.1095059

[58] Gonzalo Ramos, Matthew Boulos, and Ravin Balakrishnan. 2004. Pressure Widgets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 487–494. https://doi.org/10.1145/985692.985754

[59] Gonzalo A. Ramos and Ravin Balakrishnan. 2007. Pressure marks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>San Jose</city>, <state>California</state>, <country>USA</country>, </conf-loc>) *(CHI '07)*.

Association for Computing Machinery, New York, NY, USA, 1375–1384. https://doi.org/10.1145/1240624.1240834

[60] Christian Rendl, Patrick Greindl, Kathrin Probst, Martin Behrens, and Michael Haller. 2014. Presstures: Exploring Pressure-Sensitive Multi-Touch Gestures on Trackpads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 431–434. https://doi.org/10.1145/2556288.2557146

[61] Simon Robinson, Céline Coutrix, Jennifer Pearson, Juan Rosso, Matheus Fernandes Torquato, Laurence Nigay, and Matt Jones. 2016. Emergeables: Deformable Displays for Continuous Eyes-Free Mobile Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3793–3805. https://doi.org/10.1145/2858036.2858097

[62] Olivia G Ruston, Adwait Sharma, and Mike Fraser. 2024. SeamSleeve: Robust Arm Movement Sensing through Powered Stitching. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1134–1147. https://doi.org/10.1145/3643834.3660726

[63] Deepak Ranjan Sahoo, Kasper Hornbæk, and Sriram Subramanian. 2016. TableHop: An Actuated Fabric Display Using Transparent Electrodes. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3767–3780. https://doi.org/10.1145/2858036.2858544

[64] Rei Sakura, Changyo Han, Yahui Lyu, Keisuke Watanabe, Ryosuke Yamamura, and Yasuaki Kakehi. 2023. LattiSense: A 3D-Printable Resistive Deformation Sensor with Lattice Structures. In *Proceedings of the 8th ACM Symposium on Computational Fabrication* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 2, 14 pages. https://doi.org/10.1145/3623263.3623361

[65] Toshiki Sato, Jefferson Pardomuan, Yasushi Matoba, and Hideki Koike. 2014. ClaytricSurface: An Interactive Deformable Display with Dynamic Stiffness Control. *IEEE Computer Graphics and Applications* 34, 3 (2014), 59–67. https://doi.org/10.1109/MCG.2014.39

[66] Martin Schmitz, Sebastian Günther, Dominik Schön, and Florian Müller. 2022. Squeezy-Feely: Investigating Lateral Thumb-Index Pinching as an Input Modality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 61, 15 pages. https://doi.org/10.1145/3491102.3501981

[67] Brian Schmult and Robert Jebens. 1993. Application areas for a force-feedback joystick. In *Advances in robotics, mechatronics and haptic interfaces*. ASME, ASME, Orleans, LA, USA, 47 – 54.

[68] Kang Shi, Pourang Irani, Sean Gustafson, and Sriram Subramanian. 2008. PressureFish: a method to improve control of discrete pressure-based input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1295–1298. https://doi.org/10.1145/1357054.1357256

[69] Mike Sinclair, Michel Pahud, and Hrvoje Benko. 2013. TouchMover: Actuated 3D Touchscreen with Haptic Feedback. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces* (St. Andrews, Scotland, United Kingdom) *(ITS '13)*. Association for Computing Machinery, New York, NY, USA, 287–296. https://doi.org/10.1145/2512349.2512805

[70] Mike Sinclair, Michel Pahud, and Hrvoje Benko. 2014. TouchMover 2.0 - 3D touchscreen with force feedback and haptic texture. In *2014 IEEE Haptics Symposium (HAPTICS)*. Institute of Electrical and Electronics Engineers, Houston, Tx, USA, 1–6. https://doi.org/10.1109/HAPTICS.2014.6775425

[71] M. A. Srinivasan and R. H. LaMotte. 1995. Tactual Discrimination of Softness. *Journal of Neurophysiology* 73, 1 (Jan. 1995), 88–101. https://doi.org/10.1152/jn.1995.73.1.88

[72] Mandayam A. Srinivasan and Robert H. LaMotte. 1996. Tactual Discrimination of Softness: Abilities and Mechanisms. In *Somesthesis and the Neurobiology of the Somatosensory Cortex*, O. Franzén, R. Johansson, and L. Terenius (Eds.). Birkhäuser Basel, Basel, 123–135. https://doi.org/10.1007/978-3-0348-9016-8_11

[73] Andrew A. Stanley, James C. Gwilliam, and Allison M. Okamura. 2013. Haptic Jamming: A Deformable Geometry, Variable Stiffness Tactile Display Using Pneumatics and Particle Jamming. In *2013 World Haptics Conference (WHC)*. IEEE Computer Society, Washington, DC, USA, 25–30. https://doi.org/10.1109/WHC.2013.6548379

[74] Cameron Steer, Teodora Dinca, Crescent Jicol, Michael J Proulx, and Jason Alexander. 2023. Feel the Force, See the Force: Exploring Visual-tactile Associations of Deformable Surfaces with Colours and Shapes. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 261, 13 pages. https://doi.org/10.1145/3544548.3580830

[75] Cameron Steer, Simon Robinson, Jennifer Pearson, Deepak Sahoo, Ian Mabbett, and Matt Jones. 2018. A liquid tangible display for mobile colour mixing. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 8, 7 pages. https://doi.org/10.1145/3229434.3229461

[76] Cameron Steer, Kim Sauvé, Anika Jain, Omosunmisola Lawal, Michael J Proulx, Crescent Jicol, and Jason Alexander. 2024. Squishy, Yet Satisfying: Exploring Deformable Shapes' Cross-Modal Correspondences with Colours and Emotions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 331, 20 pages. https://doi.org/10.1145/3613904.3641952

[77] Carolin Stellmacher. 2021. Haptic-Enabled Buttons Through Adaptive Trigger Resistance. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE Computer Society, Washington, DC, USA, 201–204. https://doi.org/10.1109/VRW52623.2021.00044

[78] Carolin Stellmacher, Feri Irsanto Pujianto, Tanja Kojic, Jan-Niklas Voigt-Antons, and Johannes Schöning. 2024. Experiencing Dynamic Weight Changes in Virtual Reality Through Pseudo-Haptics and Vibrotactile Feedback. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 421, 13 pages. https://doi.org/10.1145/3613904.3642552

[79] Craig Stewart, Michael Rohs, Sven Kratz, and Georg Essl. 2010. Characteristics of Pressure-Based Input for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 801–810. https://doi.org/10.1145/1753326.1753444

[80] Martin A. Stoffel, Shinichi Nakagawa, and Holger Schielzeth. 2021. `partR2`: partitioning R^2 in generalized linear mixed models. *PeerJ* 9 (May 2021), e11414. https://doi.org/10.7717/peerj.11414

[81] Kenji Suzuki, Ryuuki Sakamoto, Daisuke Sakamoto, and Tetsuo Ono. 2018. Pressure-Sensitive Zooming-out Interfaces for One-Handed Mobile Interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3229434.3229446

[82] Faisal Taher, Jason Alexander, John Hardy, and Eduardo Velloso. 2014. An Empirical Characterization of Touch-Gesture Input-Force on Mobile Devices. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 195–204. https://doi.org/10.1145/2669485.2669515

[83] Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined gestures for elastic, deformable displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (Como, Italy) *(AVI '14)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2598153.2598184

[84] Jessica Tsimeris, Colin Dedman, Michael Broughton, and Tom Gedeon. 2013. ForceForm: A Dynamically Deformable Interactive Surface. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces (ITS '13)*. ACM, New York, NY, USA, 175–178. https://doi.org/10.1145/2512349.2512807

[85] Jessica Tsimeris, Duncan Stevenson, Tom Gedeon, and Matt Adcock. 2013. Using ForceForm, a Dynamically Deformable Interactive Surface, for Palpation Simulation in Medical Scenarios. In *Proceedings of the Second International Workshop on Smart Material Interfaces: Another Step to a Material Future (SMI '13)*. ACM, New York, NY, USA, 19–22. https://doi.org/10.1145/2534688.2534693

[86] Anke van Oosterhout and Eve Hoggan. 2020. Reshaping Interaction with Rotary Knobs: Combining Form, Feel and Function. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1973–1982. https://doi.org/10.1145/3357236.3395536

[87] Anke van Oosterhout and Eve Hoggan. 2021. Deformation Techniques for Shape Changing Interfaces. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 388, 7 pages. https://doi.org/10.1145/3411763.3451622

[88] Anke van Oosterhout, Eve Hoggan, Majken Kirkegaard Rasmussen, and Miguel Bruns. 2019. DynaKnob: Combining Haptic Force Feedback and Shape Change. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 963–974. https://doi.org/10.1145/3322276.3322321

[89] Fabrizio Vecchi, Cinzia Freschi, Silvestro Micera, Angelo M Sabatini, Paolo Dario, Rinaldo Sacchetti, et al. 2000. Experimental evaluation of two commercial force sensors for applications in biomechanics and motor control. In *5th Ann. Conf. of Int. FES*. International Functional Electrical Stimulation Society, Graz, Austria, N/A.

[90] Feng Wang and Xiangshi Ren. 2009. Empirical evaluation for finger input properties in multi-touch interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/1518701.1518864

[91] Manying Wang, Guihe Qin, Minghui Sun, Jian Zhao, and Jun Qin. 2023. An exploration of pressure input with bare finger for Mobile interaction in stationary and Mobile situations. *Multimedia Tools and Applications* 82, 17 (2023), 1–21. https://doi.org/10.1007/s11042-023-14503-0

[92] Xiyao Wang, Lonni Besançon, Mehdi Ammi, and Tobias Isenberg. 2019. Augmenting Tactile 3D Data Navigation With Pressure Sensing. *Computer Graphics Forum* 38, 3 (June 2019), 635–647. https://doi.org/10.1111/cgf.13716

[93] Yoshihiro Watanabe, Alvaro Cassinelli, Takashi Komuro, and Masatoshi Ishikawa. 2008. The deformable workspace: A membrane between real and virtual space. In *2008 3rd IEEE International Workshop on Horizontal Interactive Human Computer Systems*. IEEE Computer Society, Washington, DC, USA, 145–152. https://doi.org/10.1109/TABLETOP.2008.4660197

[94] Martin Weigel and Jürgen Steimle. 2017. DeformWear: Deformation Input on Tiny Wearable Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 28 (jun 2017), 23 pages. https://doi.org/10.1145/3090093

[95] Yannick Weiss, Steeven Villa, Albrecht Schmidt, Sven Mayer, and Florian Müller. 2023. Using Pseudo-Stiffness to Enrich the Haptic Experience in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 388, 15 pages. https://doi.org/10.1145/3544548.3581223

[96] Michael Wessely, Theophanis Tsandilas, and Wendy E. Mackay. 2016. Stretchis: Fabricating Highly Stretchable User Interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 697–704. https://doi.org/10.1145/2984511.2984521

[97] Graham Wilson, Stephen A. Brewster, Martin Halvey, Andrew Crossan, and Craig Stewart. 2011. The Effects of Walking, Feedback and Control Method on Pressure-Based Interaction. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 147–156. https://doi.org/10.1145/2037373.2037397

[98] Graham Wilson, Craig Stewart, and Stephen A. Brewster. 2010. Pressure-Based Menu Selection for Mobile Devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/1851600.1851631

[99] Graham Wilson, Craig Stewart, and Stephen A. Brewster. 2010. Pressure-based menu selection for mobile devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/1851600.1851631

[100] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. 2020. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876

[101] Yue Yang, Lei Ren, Chuang Chen, Bin Hu, Zhuoyi Zhang, Xinyan Li, Yanchen Shen, Kuangqi Zhu, Junzhe Ji, Yuyang Zhang, Yongbo Ni, Jiayi Wu, Qi Wang, Jiang Wu, Lingyun Sun, Ye Tao, and Guanyun Wang. 2024. SnapInflatables: Designing Inflatables with Snap-through Instability for Responsive Interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 342, 15 pages. https://doi.org/10.1145/3613904.3642933

[102] Lining Yao, Ryuma Niiyama, Jifei Ou, Sean Follmer, Clark Della Silva, and Hiroshi Ishii. 2013. PneUI: pneumatically actuated soft composite materials for shape changing interfaces. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/2501988.2502037

[103] Zhuoming Zhang, Jessalyn Alvina, Françoise Détienne, and Eric Lecolinet. 2022. Pulling, Pressing, and Sensing with In-Flat: Transparent Touch Overlay for Smartphones. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI 2022)*. Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. https://doi.org/10.1145/3531073.3531111

[104] Clement Zheng, Zhen Zhou Yong, Hongnan Lin, HyunJoo Oh, and Ching Chiuan Yen. 2022. Shape-Haptics: Planar & Passive Force Feedback Mechanisms for Physical Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 171, 15 pages. https://doi.org/10.1145/3491102.3501829

[105] Clement Zheng, Zhen Zhou Yong, Hongnan Lin, HyunJoo Oh, and Ching Chiuan Yen. 2022. Shape-Haptics: Planar amp; Passive Force Feedback Mechanisms for Physical Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 171, 15 pages. https://doi.org/10.1145/3491102.3501829

A APPENDIX: QUESTIONNAIRES

A.1 NASA TLX Questionnaire

Below are the six NASA TLX questions the participants completed after finishing each block in Studies 2 and 3. The participants rated their experience on a scale of 1– 20. Before the questions, the participant saw the following introductory statement: "Click on each scale at the point that best indicates your experience of this part of the task"

NASA TLX Question 1 - Mental Demand

Question: *"How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?"*

NASA TLX Question 2 - Physical Demand

Question: *"How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?"*

NASA TLX Question 3 - Temporal Demand

Question: *"How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?"*

NASA TLX Question 3 - Performance

Question: *"How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?"*

NASA TLX Question 3 - Effort

Question: *"How hard did you have to work (mentally and physically) to accomplish your level of performance?"*

NASA TLX Question 3 - Frustration

Question: *"How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?"*

A.2 Likert Scale Questionnaire

Below are the four seven-point Likert scale questions the participants completed after finishing each sample in Studies 2 and 3. The third and fourth questions were accompanied by a short context statement to provide additional clarity.

Likert Scale Question 1 - Pleasant

Question: *"How pleasant did you find this surface for making a force input"*
Options: *"Very unpleasant", "Moderately unpleasant", "Slightly unpleasant", "Neutral", "Slightly pleasant", "Moderately pleasant", "Very pleasant"*

Likert Scale Question 2 - Comfort

Question: *"How comfortable did you find this surface for making a force input"*
Options: *"Very uncomfortable", "Moderately uncomfortable", "Slightly uncomfortable", "Neutral", "Slightly comfortable", "Moderately comfortable", "Very comfortable"*

Likert Scale Question 3 - Stiffness

Context Statement: *"Stiffness is the extent to which an object resists deformation (e.g., being squashed) in response to an applied force (e.g., your finger). Compliance is the opposite of stiffness"*
Question: *"Describe the Stiffness of the Sample from 1-7, where 1 is very compliant and 7 is very stiff"*

Options: *"Very compliant", "Moderately compliant", "Slightly compliant", "Neutral", "Slightly stiff", "Moderately stiff", "Very stiff"*

Likert Scale Question 4 - Deformable

Context Statement: *"The surface of an object can vary from deformable (such as jelly, fruits, etc.) to rigid (such as a key in a piano or keyboard)."*
Question: *"Ignoring the stiffness/compliance of the objects, please describe the surface of the sample from deformable to rigid"*
Options: *"Very deformable", "Moderately deformable", "Slightly deformable", "Neutral", "Slightly rigid", "Moderately rigid", "Very rigid"*

A.3 Interview Questionnaire

After the participants finished all trials in Studies 2 and 3, they completed a short semi-structured interview. This began with ranking the seven samples, before discussing their preferences.

Interview Question 1a Please rank the seven samples which you have interacted with today in order of your preference for using them. – Participants will rank the samples by physically selecting and arranging them.

Interview Question 1b Could you please elaborate on why you selected this ranking?

Interview Question 2a Did you have a preference between the three types of surfaces rigidity you have used (Solid, Deformable and Movable, No preference)? – Participants will be shown example samples during this question.

Interview Question 2b What did you like and dislike about the three types of surface rigidity you have used and what effect did they have on your ability to perform the task? – Participants will be shown example samples during this question.

Interview Question 3a Did you have a preference between the four levels of stiffness you have used (Solid, Hard, Medium, Soft, No preference)? – Participants will be shown example samples during this question.

Interview Question 3b What did you like and dislike about the four levels of stiffness you have used and what effect did they have on your ability to perform the task? – Participants will be shown example samples during this question.

Interview Question 4 Would you like to adjust your ranking preference from question 1?

Interview Question 5 Do you have any other comments which you want to share?

B APPENDIX: TABLES

Table 5. Results of the Friedman tests, showing the effect of SCALE and SAMPLE on the NASA TLX demand and workload measures, for Study 2. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Mental Demand	*Sample*	12.06	6.00	0.06
2		*Scale*	29.01	2.00	<.001 ***
3	Physical Demand	*Sample*	16.82	6.00	0.001 **
4		*Scale*	26.25	2.00	<.001 ***
5	Temporal Demand	*Sample*	15.43	6.00	0.0172 *
6		*Scale*	21.26	2.00	<.001 ***
7	Effort	*Sample*	7.14	6.00	0.31
8		*Scale*	16.89	2.00	<.001 ***
9	Performance	*Sample*	15.11	6.00	0.0194 *
10		*Scale*	34.84	2.00	<.001 ***
11	Frustration	*Sample*	8.62	6.00	0.20
12		*Scale*	27.29	2.00	<.001 ***

Table 6. Results of the Friedman tests, showing the effect of COMPLIANCE TYPE and STIFFNESS on the NASA TLX demand and workload measures, for Study 2. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Mental Demand	Compliance Type	0.93	1.00	0.34
2		Stiffness	1.92	2.00	0.38
3	Physical Demand	Compliance Type	7.00	1.00	0.0082 **
4		Stiffness	1.67	2.00	0.43
5	Temporal Demand	Compliance Type	3.57	1.00	0.06
6		Stiffness	0.71	2.00	0.70
7	Effort	Compliance Type	1.96	1.00	0.16
8		Stiffness	1.74	2.00	0.42
9	Performance	Compliance Type	2.46	1.00	0.12
10		Stiffness	4.89	2.00	0.09
11	Frustration	Compliance Type	1.81	1.00	0.18
12		Stiffness	0.85	2.00	0.65

Table 7. Descriptive results for the NASA TLX survey, for Study 2. The medium and MAD are shown for each level in SAMPLE, SCALE, COMPLIANCE TYPE and STIFFNESS.

			Mental Demand		Physical Demand		Temporal Demand		Effort		Performance		Frustration	
Line #	Measure		Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD
1	Sample	Soft deformable	5	4.45	7	4.45	4	2.97	7	4.45	8	2.97	4	4.45
2		Medium deformable	5	3.71	7.5	5.19	5	2.97	8	4.45	8	2.97	4	4.45
3		Hard deformable	5	4.45	7.5	5.19	4	2.97	8	5.93	8	2.97	4	4.45
4		Soft movable	5.5	4.45	8.5	4.45	4	2.97	8	5.93	9	2.97	5	5.93
5		Medium movable	6	4.45	8	5.19	5	2.97	8	5.93	9	2.97	5	3.71
6		Hard movable	6	4.45	9	4.45	6	4.45	9	4.45	9	2.97	5	4.45
7		Rigid	6	4.45	9	4.45	6	2.97	9	4.45	9	2.97	5	2.97
8	Scale	20	4.5	3.71	7	4.45	4	2.97	6	4.45	7	2.97	3	2.97
9		40	5	4.45	8	4.45	5	2.97	8	4.45	9	2.97	5	4.45
10		60	7	5.93	9	4.45	6	4.45	10	4.45	10	2.97	6	4.45
11	Compliance Type	Deformable	5	4.45	7	4.45	4	2.97	7.5	5.19	8	2.97	4	4.45
12		Movable	6	4.45	9	4.45	5	2.97	8	5.93	9	2.97	5	4.45
13	Stiffness	Soft	5	4.45	7	4.45	4	2.97	8	4.45	8	2.97	4	4.45
14		Medium	5	4.45	8	5.93	5	2.97	8	4.45	8	2.97	4	4.45
15		Hard	6	4.45	8	4.45	5	3.71	8	5.93	8	2.97	4.5	3.71

Table 8. Results of the Friedman tests, showing the effect of SAMPLE on the likert measures and the sample ranking, for Study 2. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Pleasant	Sample	36.56	6.00	<.001 ***
2	Comfort	Sample	35.25	6.00	<.001 ***
3	Stiffness	Sample	92.56	6.00	<.001 ***
4	Deformablity	Sample	107.88	6.00	<.001 ***
5	Ranking	Sample	40.36	6.00	<.001 ***

Table 9. Results of the Friedman tests, showing the effect of COMPLIANCE TYPE and STIFFNESS on the likert measures and the sample ranking, for Study 2. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Pleasant	Compliance Type	3.24	1.00	0.07
2		Stiffness	12.00	2.00	0.0025 **
3	Comfort	Compliance Type	6.76	1.00	0.0093 **
4		Stiffness	9.37	2.00	0.0093 **
5	Stiffness	Compliance Type	16.33	1.00	<.001 ***
6		Stiffness	30.40	2.00	<.001 ***
7	Deformablity	Compliance Type	26.00	1.00	<.001 ***
8		Stiffness	35.48	2.00	<.001 ***
9	Ranking	Compliance Type	2.29	1.00	0.13
10		Stiffness	6.29	2.00	0.043 *

Table 10. Descriptive results for the Likert survey and sample ranking, for Study 2. The medium and MAD are shown for each level in SAMPLE, COMPLIANCE TYPE and STIFFNESS.

			Pleasant		Comfort		Stiffness		Deformablity		Rank	
Line #	Measure		Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD
1	Sample	Soft deformable	6	1.48	6	1.48	2	0.74	2	0.00	4.5	2.97
2		Medium deformable	5	1.48	5	1.48	3	1.48	3	0.00	5	2.97
3		Hard deformable	5	1.48	5	1.48	6	0.74	5	2.22	4	1.48
4		Soft movable	5	1.48	5	1.48	5	1.48	6	1.48	5.5	2.22
5		Medium movable	5	1.48	4	1.48	5	2.97	6	1.48	4	1.48
6		Hard movable	4	1.48	3	1.48	6	1.48	7	0.00	3	1.48
7		Rigid	4	1.48	3	0.74	6	1.48	7	0.00	1	0.00
8	Compliance Type	Deformable	5	1.48	5	1.48	3	1.48	3	1.48	5	2.97
9		Movable	5	1.48	4	1.48	5	2.22	6	1.48	4	1.48
10	Stiffness	Soft	5	1.48	5	1.48	2.5	0.74	2.5	2.22	5	2.97
11		Medium	5	1.48	5	1.48	3	1.48	3	1.48	5	1.48
12		Hard	4	1.48	4	1.48	6	1.48	6	1.48	3.5	2.22

Table 11. Results of the Friedman tests, showing the effect of FINGER VISIBILITY and SAMPLE on the NASA TLX demand and workload measures, for Study 3. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Mental Demand	Sample	2.77	6.00	0.84
2		Finger Vis	0.33	1.00	0.56
3	Physical Demand	Sample	5.72	6.00	0.46
4		Finger Vis	1.81	1.00	0.18
5	Temporal Demand	Sample	3.48	6.00	0.75
6		Finger Vis	0.05	1.00	0.83
7	Effort	Sample	7.71	6.00	0.26
8		Finger Vis	1.50	1.00	0.22
9	Performance	Sample	10.39	6.00	0.11
10		Finger Vis	1.96	1.00	0.16
11	Frustration	Sample	18.97	6.00	0.0042 **
12		Finger Vis	1.50	1.00	0.22

Table 12. Results of the Friedman tests, showing the effect of COMPLIANCE TYPE and STIFFNESS on the NASA TLX demand and workload measures, for Study 3. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Mental Demand	Compliance Type	0.14	1.00	0.71
2		Stiffness	1.69	2.00	0.43
3	Physical Demand	Compliance Type	0.33	1.00	0.56
4		Stiffness	4.56	2.00	0.10
5	Temporal Demand	Compliance Type	0.18	1.00	0.67
6		Stiffness	0.88	2.00	0.64
7	Effort	Compliance Type	0.00	1.00	1.00
8		Stiffness	1.87	2.00	0.39
9	Performance	Compliance Type	0.36	1.00	0.55
10		Stiffness	1.38	2.00	0.50
11	Frustration	Compliance Type	0.15	1.00	0.69
12		Stiffness	3.86	2.00	0.14

Table 13. Descriptive results for the NASA TLX survey, for Study 3. The medium and MAD are shown for each level in SAMPLE, FINGER VISIBILITY, COMPLIANCE TYPE and STIFFNESS.

			Mental Demand		Physical Demand		Temporal Demand		Effort		Performance		Frustration	
Line #	Measure		Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD
1		Soft deformable	10	5.93	9	4.45	6	4.45	9	4.45	8.5	5.19	8	5.19
2		Medium deformable	10	4.45	9	5.93	6	3.71	10.5	3.71	10	4.45	10.5	3.71
3		Hard deformable	10	5.93	10	5.93	6.5	3.71	11	4.45	8	4.45	8	4.45
4	Sample	Soft movable	9	4.45	10	5.93	7	4.45	11	4.45	7	4.45	8	4.45
5		Medium movable	10	4.45	9.5	4.45	6.5	3.71	10	4.45	8	3.71	7	5.19
6		Hard movable	9	5.93	11	5.93	7	2.97	11.5	5.19	8	4.45	9	5.93
7		Rigid	10	4.45	11	4.45	6	4.45	12	2.97	7.5	5.19	9	4.45
8	Finger	Seen	10	4.45	10	5.93	6.5	3.71	11	4.45	8	4.45	9	4.45
10	Visibility	Unseen	10	5.93	9	5.19	6	4.45	10	4.45	8	4.45	8	4.45
11	Compliance	Deformable	10	5.93	9	5.93	6	4.45	10	4.45	9	4.45	9	4.45
12	Type	Movable	9	4.45	10	4.45	7	3.71	11	4.45	8	4.45	8	4.45
13		Soft	10	5.93	9	5.19	6	2.97	9.5	5.19	7.5	5.19	8	4.45
14	Stiffness	Medium	10	4.45	9	4.45	6	4.45	10	4.45	9	4.45	9	4.45
15		Hard	10	5.93	10	5.93	7	2.97	11	4.45	8	4.45	9	5.93

Table 14. Results of the Friedman tests, showing the effect of SAMPLE on the likert measures and the sample ranking, for Study 3. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Pleasant	Sample	33.16	6.0	<.001 ***
2	Comfort	Sample	32.09	6.00	<.001 ***
3	Stiffness	Sample	74.57	6.00	<.001 ***
4	Deformablity	Sample	100.03	6.0	<.001 ***
5	Ranking	Sample	36.92	6.00	<.001 ***

Table 15. Results of the Friedman tests, showing the effect of COMPLIANCE TYPE and STIFFNESS on the likert measures and the sample ranking, for Study 3. * indicates significant results where $p < .05$, ** significant results where $p < .01$, *** and significant results where $p < .001$

Line #	Measure	Effect	χ^2	DFn	p
1	Pleasant	Compliance Type	7.35	1.00	0.0067 **
2		Stiffness	13.55	2.00	0.0011 **
3	Comfort	Compliance Type	5.26	1.00	0.022 *
4		Stiffness	10.51	2.00	0.0052 **
5	Stiffness	Compliance Type	1.96	1.00	0.162
6		Stiffness	41.22	2.00	<.001 ***
7	Deformablity	Compliance Type	17.64	1.00	<.001 ***
8		Stiffness	32.76	2.00	<.001 ***
9	Ranking	Compliance Type	0.04	1.00	0.847
10		Stiffness	14.61	2.00	<.001 ***

Table 16. Descriptive results for the Likert survey and sample ranking, for Study 3. The medium and MAD are shown for each level in SAMPLE, COMPLIANCE TYPE and STIFFNESS.

			Pleasant		Comfort		Stiffness		Deformablity		Rank	
Line #	Measure		Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD	Medium	MAD
1		Soft deformable	6.0	1.5	6.0	1.5	2.0	1.5	2.0	1.5	5.0	3.0
2		Medium deformable	5.0	1.5	5.0	1.5	3.0	1.5	3.0	1.5	5.0	3.0
3		Hard deformable	4.0	1.5	4.0	1.5	6.0	1.5	5.5	0.7	3.0	1.5
4	Sample	Soft movable	5.0	1.5	5.0	1.5	3.5	2.2	5.0	1.5	5.0	3.0
5		Medium movable	5.0	1.5	5.0	1.5	5.0	2.2	5.0	1.5	5.0	3.0
6		Hard movable	4.0	1.5	4.0	1.5	6.0	1.5	6.0	1.5	3.0	1.5
7		Rigid	4.5	2.2	3.5	1.5	6.0	1.5	6.5	0.7	1.0	0.0
8	Compliance	Deformable	5.5	2.22	5	1.48	4	2.97	3	1.48	3	2.97
9	Type	Movable	5	1.48	4.5	2.22	5	2.22	5	1.48	3	1.48
10		Soft	6	1.48	5.5	2.22	3	1.48	3	2.97	5	2.97
11	Stiffness	Medium	5	1.48	5	1.48	3	1.48	3.5	2.22	5	2.97
12		Hard	4	1.48	4	1.48	6	1.48	6	1.48	3	1.48