Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device

Sean Gustafson, Christian Holz and Patrick Baudisch
Hasso Plattner Institute
Potsdam, Germany
{sean.gustafson, christian.holz, patrick.baudisch} @ hpi.uni-potsdam.de

ABSTRACT

We propose a method for learning how to use an imaginary interface (i.e., a spatial non-visual interface) that we call "transfer learning". By using a physical device (e.g. an iPhone) a user inadvertently learns the interface and can then transfer that knowledge to an imaginary interface. We illustrate this concept with our *Imaginary Phone* prototype. With it users interact by mimicking the use of a physical iPhone by tapping and sliding on their empty non-dominant hand without visual feedback. Pointing on the hand is tracked using a depth camera and touch events are sent wirelessly to an actual iPhone, where they invoke the corresponding actions. Our prototype allows the user to perform everyday task such as picking up a phone call or launching the timer app and setting an alarm. Imaginary Phone thereby serves as a shortcut that frees users from the necessity of retrieving the actual physical device.

We present two user studies that validate the three assumptions underlying the transfer learning method. (1) Users build up spatial memory automatically while using a physical device: participants knew the correct location of 68% of their own iPhone home screen apps by heart. (2) Spatial memory transfers from a physical to an imaginary interface: participants recalled 61% of their home screen apps when recalling app location on the palm of their hand. (3) Palm interaction is precise enough to operate a typical mobile phone: Participants could reliably acquire 0.95cm wide iPhone targets on their palm—sufficiently large to operate any iPhone standard widget.

Author Keywords

Imaginary interface, mobile, wearable, spatial memory, screen-less, memory, non-visual, touch.

ACM Classification

H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms

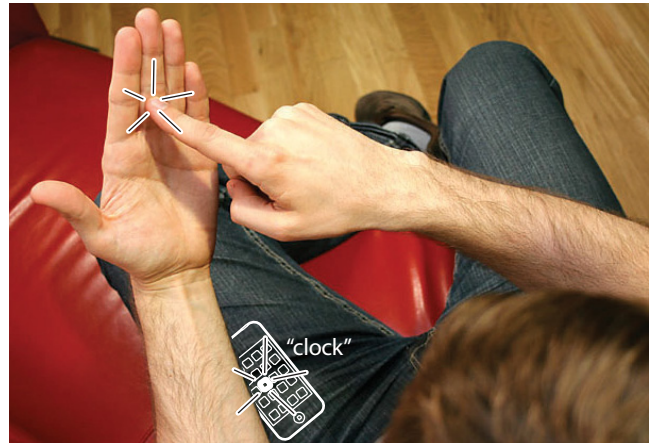Design**,** Human Factors, Experimentation.

Figure 1: This user operates his mobile phone in his pocket by *mimicking* the interaction on the palm of his non-dominant hand. The palm becomes an *Imaginary Phone* that can be used in place of the actual phone. The interaction is tracked and sent to the actual physical device where it triggers the corresponding function. The user thus leverages spatial memory built up while using the screen device. We call this *transfer learning*.

INTRODUCTION

Imaginary interfaces were proposed as a means for enabling pointing input on screen-less mobile devices [5]. With their hands tracked by a chest-worn camera, users of imaginary interfaces point and draw in the *empty space* in front of them. Their non-dominant hand, held up in an "L-gesture", forms the origin of a 2D coordinate system. This visual reference allows users to acquire targets using coordinates of the style "two thumbs up and three index fingers to the right". This allowed for reliable acquisition of targets measuring 4.8×4.3cm.

However, if we try to transfer this approach to *multi-widget* imaginary interfaces, we obtain an interaction style reminiscent of a voice menu: the system would have to read out choices such as "For mail, select one thumb right, two index fingers up. For weather…" Having to listen to such a list makes interaction slow and frustrating [26]. Extended use would eventually allow users to select widgets without listening to the choices anymore (to *dial ahead* [18]), but since real-world interfaces can hold dozens of widgets, learning all the widget locations can take a long time, leaving users stuck with the voice-menu style of interaction.

Transfer learning

We propose an alternative approach which we call *transfer learning*. By designing imaginary interfaces that mimic the layout of a mobile device that users are already familiar with, users are able to operate an imaginary interface by *mimicking* their use of the corresponding real-world screen interface (Figure 1). As we demonstrate in this paper, this allows users to apply the spatial knowledge gathered on the physical device to an imaginary interface. We will refer to this by saying that the spatial knowledge *transfers* from the physical device to the imaginary interface.

CONTRIBUTION

This paper has two main contributions: (1) the concept of learning imaginary interfaces by *transfer* and as an example of that concept (2) our *Imaginary Phone* prototype. We also present two users studies that serve to validate the underlying principles of transfer learning and the Imaginary Phone prototype.

THE IMAGINARY PHONE

We start by illustrating the concept with a prototype we call Imaginary Phone: an imaginary interface that offers a shortcut interface for an iPhone (Figure 2). Instead of retrieving and operating the physical phone, users *mimic* the interaction by pointing and dragging on their empty hand. Our prototype tracks the pointing interaction between the two hands (see *Prototype and Tracking Hardware*) and sends the touch position to a physical mobile device, here an iPhone located in the user's pocket. The physical device supplies feedback to operations via the built-in speaker or a wireless headset worn by the user.

Walkthrough

Users can choose, either because it is necessary or just convenient, to use their Imaginary Phone for various quick tasks instead of retrieving the physical device from their pocket. Here is an example scenario:
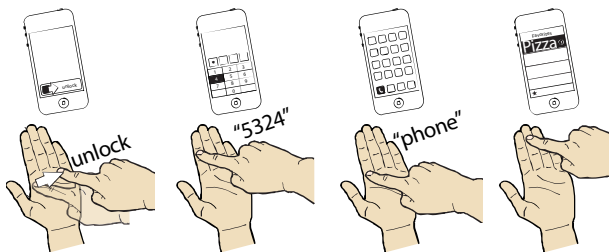

Figure 2: Walkthrough of making a call with the Imaginary Phone: (a) unlock with a swipe, (b) enter your pin, (c) select the 'phone' function and (d) select the first entry from the speed dial list.

Karl is cleaning up the dishes and receives a phone call. Since his hands are wet, he cannot take the call on his physical phone and uses the Imaginary Phone instead. He answers the call by swiping on his hand, which is the same interaction he would have performed on the physical phone. The call is from a friend that wants to go jogging

tomorrow morning. Karl ends the call by touching the location of the 'End' button on this wet palm then launches the 'Clock' application. From here he selects the location for "Alarms" and enables his early morning alarm to ensure he gets up in time.

Later, while watching TV, Karl wants to order food but cannot find his phone in his pockets. Not wanting to get up from the couch and search, he chooses the Imaginary Phone to place a call as shown in Figure 2.

RESULTING INTERACTION MODEL AND BENEFITS

Interaction with an imaginary interface that is learned by transfer, such as by Karl in the previous scenario, is possible only because the user has been using the physical screen device over a period of time and has learned the spatial locations of the necessary user interface elements. This happened inadvertently—without extra effort users become increasingly familiar with the locations of such widgets over time. The spatial knowledge they gained from using the physical device can be *transferred* to an imaginary interface.

At some point, the user has performed an operation often enough to know the locations and sequence of touches needed to execute it and he can begin to perform that operation on the imaginary version of the interface. This will occur one operation at a time, with the simpler and more common operations being transferred earlier. Therefore, microinteractions [3] will generally be the first to transition to an imaginary interface.

As a result, this transfer model essentially turns the screen device into a *training mode* for the imaginary interface—or, depending on your perspective, the imaginary interface into an *expert mode* for the screen device. Accordingly, the benefit of the transfer model depends on the use case:

For users of physical devices, the main benefit of the transfer model is that it allows mobile phone users to instead perform their interactions on an imaginary interface. This saves these users the effort of retrieving the physical device, because the shorter the interaction, the greater the relative speedup. The latter makes the transfer model particularly valuable for microinteractions, such as dismissing an alarm dialog. Since the transfer model allows users to leverage their experience with the physical device, users can redeem these benefits right away, without the need for a separate training period.

For users of imaginary interfaces (i.e., users that only have access to the imaginary interface), the transfer model replaces the voice menu-style training period. Offloading the learning phase to a screen device, (1) allows learning to take place in a visual and inherently parallel way and, (2) unlike when using a voice menu-style interface, interaction is fast during training, lowering the entrance barrier to learning imaginary interfaces.

PROTOTYPE AND TRACKING HARDWARE

Our prototype senses touch using a depth camera and, after processing, injects touch events into a standard iPhone.

Sensing hardware

While imaginary interfaces as originally proposed required an infrared camera [5], interaction on the palm opens up other sensing options, such as gloves [11], Skinput [6] or depth cameras. We chose to use a time-of-flight depth camera, because unlike other approaches, it supports interaction with empty hands and allows for dragging interaction. Although currently mounted on a tripod and looking over the user's shoulder, we imagine future depth cameras to be small enough to be mounted on the chest as originally proposed for imaginary interfaces [5].

Our choice to use a time-of-flight camera can lead to occlusion issues (e.g., back of the hand occluding the pointing finger), but it does enable our prototype to work in all lighting conditions, including outside in direct sunlight (as shown in Figure 3b-c), unlike standard infrared cameras or Kinect [14]. Our depth camera is a PMD[vision] CamCube that provides frames at 40Hz with 200×200px resolution.

Figure 3: (a) We track input using a time-of-flight depth camera (PMD[vision] CamCube), which allows our Imaginary Phone prototype to also work in direct sunlight (b-c).

Algorithm

In order to extract the two hands from the input image, we pre-process the raw depth image as shown in Figure 4. We first find the closest pixels in the depth image (a), remove all pixels with relative depth values of more than 30cm, and smooth all remaining values. To determine the number of visible hands, we create a depth histogram of the masked image (c) and calculate the number of strong peaks (indicated by green squares in Figure 4c). Based on the two distributions in the histogram, we classify pixels in the depth image (d) to obtain the masks for the pointing hand (e) and the reference hand's palm (f).

To determine if and where the user is touching the palm, we pick a location inside the pointing hand (Figure 4e) and fill using a small tolerance value, eventually walking "down" the finger towards the reference hand (f). If the fill does not reach a depth value that belongs to the reference hand while staying within the tolerance value, we infer no touch. If it does, we infer that the finger is touching.

Due to the limited resolution of the depth camera, we cannot find the precise end of the touching finger. Instead, we determine the touch location from the end of the point mask offset by a small vector in the direction of the finger (green square in Figure 4g).

The width of the reference frame for touch events is set to the width of the fingers excluding the thumb (see *Design Discussion* for other possibilities). We first calculate the width from the top 3cm of the hand to exclude the thumb (the depth values allow translating this into pixels measurements) and draw a frame around those values. We then set the height of this reference frame to match an aspect ratio of 1.5. The final frame is shown in Figure 4f and g. As this reference frame is subject to noise if the pointing hand is present, we update the reference frame only if one hand is visible and, upon sensing both hands, adapt it by tracking the reference hand.
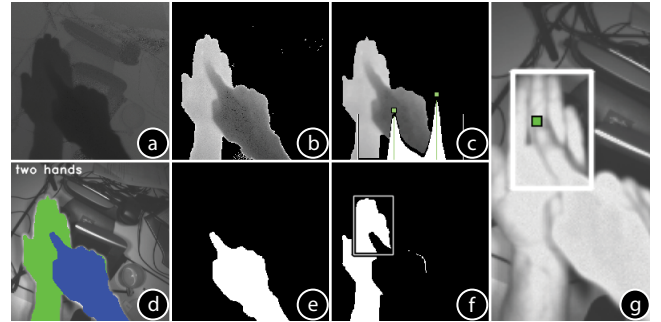

Figure 4: In processing the (a) raw depth image, our system (b) thresholds and (c) calculates a depth histogram to (d) segment the image into two masks: (e) pointing hand and (f) reference hand. From that we calculate (g) the final touch position and reference frame.

As the computed raw locations are subject to strong noise, we use hysteresis to maintain touch states (touch/no touch) and smooth input coordinates, which enables smooth dragging or even free-form drawing. This also prevents processing inadvertent input, such as a hand waving by the camera. Our system supports all of the same single-touch interactions that are possible on the phone: swiping, scrolling, tapping, dragging, drawing, etc.

After determining touch position on the palm, our prototype relays touch input to the iPhone. A custom-written input daemon on the iPhone receives the smoothed events via TUIO over WiFi and injects them into the event stream of the iPhone. The VoiceOver [1] accessibility mode (built into Apple iOS 4.0 and greater) provides auditory confirmation of actions. The unlock gesture on the iPhone to prevent inadvertent touch input additionally helps our system to disregard spurious input through gestures that happen naturally when not using the system.

RELATED WORK

Imaginary Phone draws on several areas of related work: wearable and mobile computing, microinteractions, sensing approaches for using the hand as an input surface and imaginary interfaces.

Wearable and mobile computing

Like Imaginary Phone there is a wide range of wearable and mobile computing systems that allow the user to interact without holding a device in their hands. Notably,

BodySpace [20] assigns functions to position on the body, Abracadabra [7] senses movement of a magnet placed on the finger tip from a wrist-mounted display and GesturePad [19] provides a touch sensitive pad embedded into clothing.

Microinteractions

Imaginary Phone is particularly well suited for supporting microinteractions—the quick mobile device interactions that characterize the dominant interaction mode for mobile phones [3]. Ashbrook et al. [2] showed that it takes over 4.5 seconds on average just to begin an interaction with a mobile phone stored in your pocket. This is a substantial overhead for an interaction that overall only lasts a few seconds [17].

Using the hand as input surface

Unlike the original imaginary interfaces, this paper proposes using the palm of the non-dominant hand as an interaction surface. Several projects, such as Haptic Hand [10], Sixth Sense [15] and Brainy Hand [21], also propose using the palm.

There are a few technical approaches to sensing touch on the surface of the hand. First, a designer could place sensing material on the palm. For example, KITTY [11] covers parts of the hand with electrical contacts that, when touched with another contact, register a touch event at a specific location. Although this method could produce highly reliable and perhaps high-resolution input, the fact that the user must wear something over their hand prohibits the general use of this approach.

Second, a system could observe the physical manifestations of touch from afar. For example, GestureWrist [19] senses hand postures by observing the changes in the shape of the wrist and Skinput [6] senses taps on the hand and forearm by measuring the different patterns of vibrations that travel up the arm.

Finally, computer vision has been used for a wide range of hand sensing [23] but we are unaware of any computer vision based approach that allows the user to use their uninstrumented finger to select a position on their bare hand. This, perhaps, has not been done previously because of the difficulties of separating the two hands. Using a depth camera allowed us to segment both hands and sense touch on the surface of the palm.

Imaginary interfaces

Imaginary interfaces are spatial, non-visual interfaces [5]. They were originally proposed as free-hand interface based on a finger and thumb coordinate system as shown in Figure 5.

Other researchers have proposed interfaces that are non-visual for controlling input (e.g., Mouse-less [16], Virtual Shelves [12]) or describing shapes and objects (e.g., Spatial Sketch [25], Data Miming [9]).
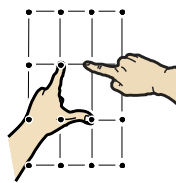
Figure 5: The finger and thumb coordinate system for the original imaginary interfaces [5].

MAKING IT WORK: THE THREE REQUIREMENTS

The Imaginary Phone is based on the concept of transfer learning that can be broken down into a chain of three logical steps (Figure 6), each of which depends on one assumption:

1) Spatial memory: while using a screen device, users inadvertently learn where user interface elements are located.

2) Transfer: with an appropriate mapping, spatial memory acquired on a physical device can be recalled on an imaginary interface.

3) Accuracy: the imaginary interface allows users to point with sufficient accuracy to provide the pointing accuracy required by the associated physical device.

(2) transfer spatial knowledge from device to palm

(1) build spatial memory from physical device

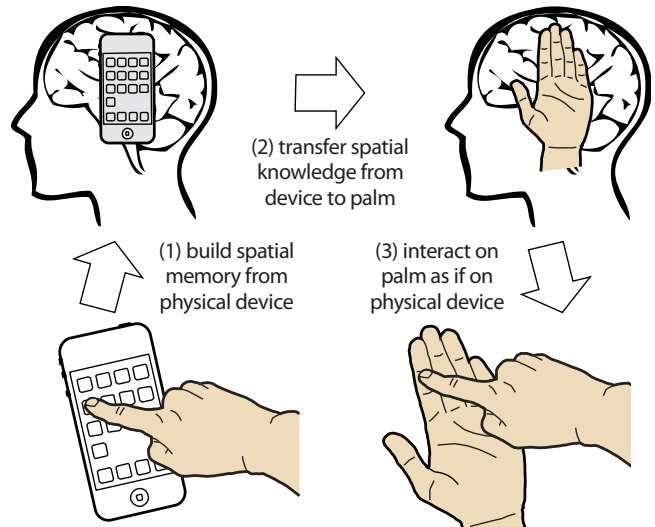(3) interact on palm as if on physical device

Figure 6: Our design is based on three assumptions: (1) using a physical device builds spatial memory, (2) the spatial memory transfers to the imaginary interface and (3) users can operate the imaginary interface with the accuracy required by the physical device.

These three assumptions inform the design of transfer-based imaginary interfaces and in particular Imaginary Phone. Next, we support these assumptions with a design discussion and empirical results from two studies.

DESIGN DISCUSSION

Assumption 2 (transfer) and 3 (accuracy) depend heavily on the design of the imaginary interface. Here we present the design alternatives we explored and explain the rationale for our choices.

Pointing in empty space vs. on the palm

The original imaginary interface lets users point in empty space, framed by an L-gesture (Figure 7a). When designing Imaginary Phone, we moved the interaction to the palm of the non-dominant hand (Figure 7b).

While the original imaginary interfaces concept, based on empty space, offered a much larger interaction area, the palm-based version offers a benefit we consider essential for transfer learning: *memorable landmarks*. The findings in [5] indicate that proximity to landmarks—in that case the

tip of index finger and thumb—helps acquire targets; yet, the empty space design is all but void of landmarks. The palm, in contrast, is full of landmarks, many of which even have commonly known names, allowing users to create symbolic associations. The obvious match between four fingers and four-column layouts in, for example, the home screen, made it clear that the palm was the better platform for Imaginary Phone.
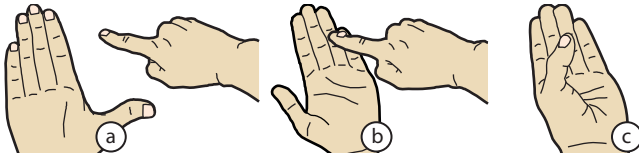

Figure 7: (a) The original Imaginary Interface had users interact in empty space framed by the L-gesture. (b) In this paper, we moved the interaction onto the non-dominant hand, which conceptually also allows (c) one-hand interaction.

As a side effect, on the palm, a tap is established by the physical contact, very much like on any touch screen. This results in four additional benefits:

1. Stabilize finger: physical contact between hands stabilizes the finger during pointing.

2. Eliminate pinching: most users are more experienced and thus skilled with tapping than with the pinching gesture required by the original empty space version.

3. Spatial haptic feedback: during tapping, the sensation on the non-dominant hand reflects the acquired location, providing an additional cue for target location.

4. Eliminate parallax: when targeting in empty space, the finger is free to move in 3D and is thus often outside the 2D interaction plane of the imaginary interface. Mapping the finger position to the desired 2D point on the plane, however, is subject to ambiguity and pointing error because we cannot know how the user conceptualizes this projection: orthogonal projection, line-of-sight, etc. Pointing on the palm avoids this problem.

While it is hard to compare the overall number of addressable locations in empty space to the palm (with empty space, pointing resolution decreases with distance from the reference hand [5]) the combination of the four factors listed above increases the pointing resolution on the palm to the point where it provides the resolution required to address widgets on an iPhone (see User Study 2). Also, the size of the palm matched the physical device much better than the large low-resolution empty space, whose large size could prove to be socially awkward.

Finally, interaction on the palm also allows for a one-handed version (as in PinchWatch [13]), illustrated in Figure 7c. This would allow for even more immediate use, albeit with less interaction space.

Defining the mapping from palm to physical screen
Palm and device screen generally do not have the same size and shape, requiring us to define a *mapping*.

Our prototype uses a simple regular grid mapping as illustrated by Figure 8b. This layout allows users to simply imagine the bounding rectangle of their hand and use that to find the position. Even more importantly, the layout is generic, thus applies to any interface including free form input, such as sketching and handwriting.
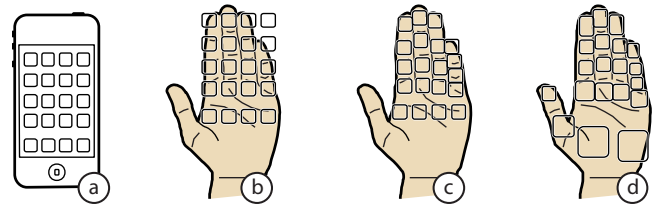

Figure 8: Mapping (a) the iPhone home screen mapped to (b) a regular grid, (c) a semi-regular grid where the columns are mapped to fingers and (d) arbitrary mapping to the best landmarks.

The more specific layouts (Figure 8c and d) should allow for increased pointing accuracy by making even better use of landmarks, but could cause confusion when trying to operate controls that assumed a rectilinear screen (Figure 9). Highly specialized layouts (Figure 8d) are impractical, as they require users to relearn mappings on a per-application basis.
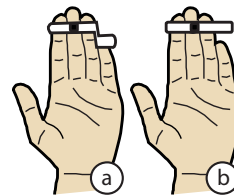

Figure 9: (a) Non-regular mappings fail when placing sliders and list items that span the width of the screen. (b) The regular grid works fine.

Our Imaginary Phone prototype uses the "4-finger scale" as shown in Figure 10c. This maps input from the larger palm to a smaller screen size, resulting in a scale factor of approximately 1.86. Unlike the more obvious 1:1 mapping (Figure 10b) the scaling allows us to include additional landmarks and thus increase the effective pointing accuracy on the mobile device. We could continue this logic by increasing the scale to include the whole hand but at the expense of leaving a large amount of the interaction space off the surface of the hand—a space devoid of landmarks.
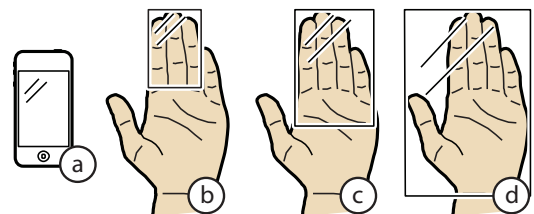

Figure 10: (a) The screen on a current 5 × 7cm mobile devices (b) maps to approximately three fingers of an adult male hand. (c) Using a scaled mapping allows us to map to four finger or (d) the whole hand. The iPhone and hands are to scale.

The four-finger scale works best with interfaces laid out in four column layouts, such as the iPhone home screen (Figure 11a). Other layouts, such as a seven-column month

calendar (Figure 11b) could be mapped by assigning every other day to the space between two fingers, which also make good landmarks. Similarly, we could map three-column layouts (Figure 11c) to only the spaces between fingers.

Figure 11: iOS screens laid out in (a) four, (b) seven, (c) and three column grids.

Some mappings, such as the semi-regular grid (Figure 8c) are simple enough to communicate with a diagram (as we did in Study 1) but with others, such as the regular grid (Figure 8b), it is not clear how user interface elements map to specific features on the user's hand. Figure 12 shows an approach we explored to teach users the regular grid mapping. If we have access to the device's wallpaper, we display a photo of the user's hand as wallpaper. During use users now learn not only target locations but also the mapping from a widget to its location on the user's palm. This design allowed us to minimize learning time in Study 2.

Figure 12: A photo of the user's hand as wallpaper helps learn the association between widget and location on the user's palm.

The arguments just presented form logical support for the three assumptions that transfer learning is based on. The following two studies complement this discussion with empirical results.

STUDY 1: RECALL YOUR HOME SCREEN LAYOUT AND TRANSFER IT TO YOUR HAND

This first study investigated the first two of the three assumptions behind transfer learning as illustrated in Figure 6. First, we wanted to know how much spatial memory users build up through the regular use of a touch-screen mobile device. Second, we wanted to know how much of that knowledge transfers to the hand. Together, these numbers would tell us *how useful* an Imaginary Phone could be. We did not test a specific hypothesis in this study. We were purely interested in participants' recall abilities.

To investigate this we asked daily iPhone users to recall the locations of the (up to) twenty home screen app icons of their own iPhone from memory and without feedback. In a

between-subjects design, half of the participants recalled and communicated their choice by pointing to a non-functional iPhone prop (*phone prop* condition Figure 13a) while the other half recalled locations by pointing to the palm of their own non-dominant hand, using a predefined scheme of how buttons on their iPhone would map to locations on their palm (*palm* condition, Figure 13b).

Research questions
The goal of this study was to determine how many app locations had been learned as a side effect of regular use and how much of that knowledge would successfully transfer, using the supplied mapping, to the participants' palm. Participants in the *palm* condition would not only have to recall, but also map locations onto their hand. The difference between the two conditions would serve as an indication for how much information is lost in transfer. We also expected the frequency of use to correlate with the user's ability to recall.

Task and procedure
After we seated participants, they unlocked their phone and, without looking at the screen, handed it over to us. Participants in the *palm* condition were now taught the semi-regular-grid mapping scheme (see Figure 8c in the Design Discussion). This preparation took less than a minute for all participants.

The experimenter then conducted a series of trials, one for each app on the participant's home screen. For each trial, the experimenter picked a different app and cued the participant with the app's name and a description of the app icon's visual appearance.

Participants responded by pointing to the app's presumed location within the 4×5 icon home screen. Participants in the *phone prop* condition pointed to cells displayed on a printed prop of an iPhone (unlabeled all-white icons in Figure 13a). Participants in the *palm* condition instead pointed to a location on their own non-dominant hand (Figure 13b).

In both conditions, the experimenter determined what location the participant was pointing to by observing them point. While we did not measure pointing accuracy directly (we investigated that in Study 2), the experimenter had no difficulty identifying which targets participants referred to.
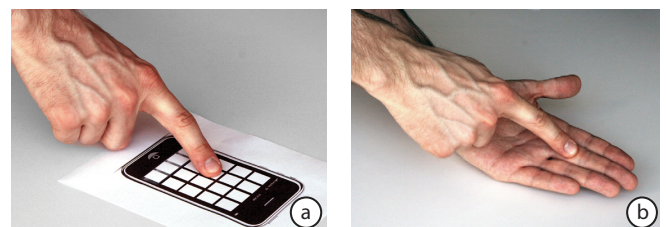

Figure 13: Study 1 task: (a) participants in the *phone prop* condition recalled app locations by pointing to an empty iPhone prop, (b) participants in the *palm* condition pointed on their own non dominant hand.

After completing all trials, participants classified each of their home screen apps as used either *daily* (at least once a day), *weekly* (at least once a week) or *rarely* (less than once per week).

Finally, participants filled out a demographic questionnaire. All participants completed the study in 15 minutes or less.
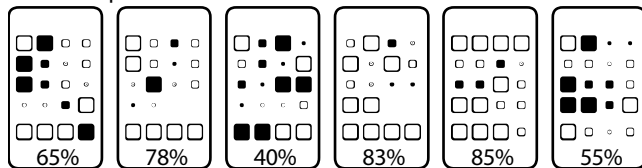
Participants
We recruited 12 participants (5 female) in the cafeteria of our institution. Participants were on average 23.6 years old (SD=4.2) and two were left-handed. All participants were daily iPhone users and carried it with them. They were given a small gift for their time.
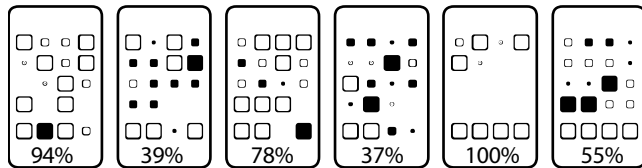
Results
The twelve participants had on average 18.4 apps on their home screens and recalled each only once, for a total of 221 app recall trials. No outliers were removed but three trials were discarded (leaving 218 for analysis) because of errors by the experimenter. Figure 14 shows the responses from all participants.

Phone Prop Selection

Palm Selection

Figure 14: Study 1 raw results: each of the 12 rounded rectangles represents one participant's phone home screen. Each black (wrong) or white (correct) square represents a specific home screen app. Percentages indicate the participant's recall rate.

Our main finding was that participants, on average, correctly positioned 64% of the apps on their phone (68% for *phone prop*, 61% for *palm*). The success rate was higher for apps used daily (71% for *phone prop*, 80% for *palm*). T-tests did not show any significant differences. When they were wrong, 45% of guesses were only a single cell off, suggesting that participants had some spatial knowledge. Figure 15 shows these aggregated statistics.

Overall the frequency of use of an application correlated with percentage correct (Pearson's r_3=0.998, p=0.043) but we found no trends relating performance to age, gender or duration of phone ownership.
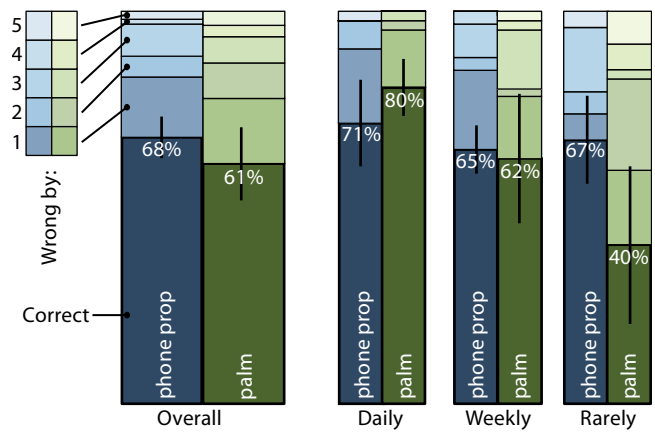
Figure 15: Study 1 aggregated results: percentage correct by use frequency (+/- std. error of the mean). The chart is stacked with mean percentages for incorrect responses separated by how far (in Manhattan distance) they were wrong by.

Discussion
Since none of the participants were aware of the task or project before the study, a mean recall rate of 64% of their home screen apps can only be explained as a side effect of regular phone use. This supports the first of our three main assumptions behind the transfer learning approach.

Note that the recall rates observed with these untrained users effectively form a lower bound as actual users of an Imaginary Phone would have an incentive to actively learn locations.

We did not find a significant difference between recall on *phone prop* and *palm* conditions. However, while the lack of significance is expected given the small number of participants and high variation, the fact that both numbers are in the same range suggests that the loss of spatial knowledge during transfer cannot be too large. This is also supported by our observations—participants seemed to recall on their hands almost as easily as on a phone. This supports the second of our three main assumptions: spatial knowledge can indeed transfer to the hand.

STUDY 2: TARGETING ON ARBITRARY LAYOUTS
Our goal of this study was to verify our design's third main assumption—that a palm-based imaginary interface, such as Imaginary Phone, supports sufficient pointing accuracy to operate the typical functions of a mobile phone. In particular, we wanted to know whether this interaction style would allow users to operate the widget sizes common to today's touch devices. If so, imaginary interfaces that mimic touch devices would become viable. To be able to compare palm interaction to previous work [5] (that used a different sensing mechanism) we included the traditional empty space imaginary interface (Figure 16a) as a control.

In order to evaluate the *concept* rather than the current condition of our Imaginary Phone prototype (whose touch resolution is limited by the depth camera's 200 × 200px resolution) we conducted this study using "perfect" tracking, i.e., post-hoc analysis of high resolution photos.

In contrast to Study 1, where targets were directly associated with hand landmarks, we chose to conduct this experiment with randomly placed targets. Therefore these results show performance worse than with a layout designed to align widgets with landmarks on the hand.

(Imaginary) interfaces tested in this study

In the *empty space* control condition participants targeted in the space framed by their thumb and index finger (Figure 16a) and in the *palm* condition participants targeted on palm of their hand (Figure 16b). The size of the tracked area was kept constant for both conditions.

To shorten training, we used the *wallpaper* approach described earlier in the design section (see Figure 12) where participants' own hand was displayed behind the targets. With this approach, they were able to readily associate the arbitrary target locations with landmarks on their own hand.
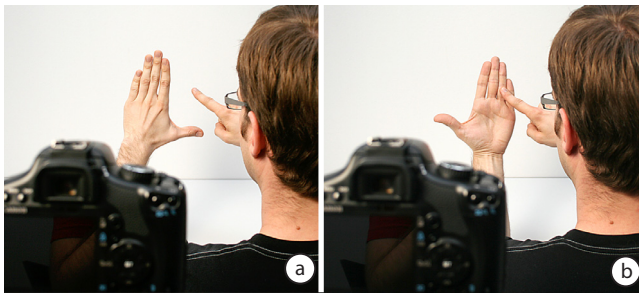

Figure 16: Study 2 apparatus: (a) *empty space* condition and (b) *palm* condition.

Task and procedure

During each trial, participants targeted three locations at a time (in pilots we determined participants were able to remember three locations easily). Participants learned the three target locations by repeatedly targeting them on a screen device (here an iPod Touch) until they were able to reliably target with at least 5mm accuracy on the touch screen. Participants were then prompted repeatedly with a target number and responded by recalling the respective position in *empty space* or on their *palm* depending on the condition.

There were two independent variables: Target Location (4 groups of 3 locations) and Interface (*empty space* vs. *palm*). As a within-subjects design, half of the blocks used the *empty space* interface and the other half the *palm* interface. The order of interfaces was counterbalanced across participants. Participants recalled and touched each target five times in random order; each participant completed four such blocks (two for each interface) with each block featuring a different set of 3 targets. Together, the experiment consisted of 12 participants × 2 interfaces × 2 blocks × 3 targets × 5 reps = 720 trials. Each participant completed the experiment session within 30 minutes.

Apparatus

We used a DSLR camera to record participants' touch interactions (Figure 16). We extracted touch locations from the high-resolution photos on a millimeter level, which kept tracking errors to a minimum.

Participants

12 participants (1 female) were recruited at our institution. They were between the ages of 19 and 28 (M=21.8, SD=2.56). All participants were right-handed. They were given a small gift for their time.

Research question and hypothesis

The main purpose of this study was to determine the accuracy of the *palm* interface. If we should find that the minimum button size of the hand interface is comparable to studies of conventional touch (such as the 15mm buttons from [8]) that would indicate that the transfer concept was viable.

Reflecting our earlier discussion on the properties of the palm interface, we hypothesized that the *palm* interface would allow participants to target with higher accuracy than in the *empty space* condition.

Results

We discarded 7 bad trials where no data was recorded and 15 outlier trials where the touch location was greater than 3 standard deviations away from the centroid. This left us with 720-7-15=698 trials in our analysis. To allow for comparison between participants, we normalized all hand sizes so that the index finger was 7.25cm long (the population's average index finger length [4]). This length corresponds to 3.90cm in the iPhone, leaving us with a scaling factor of 1.86.
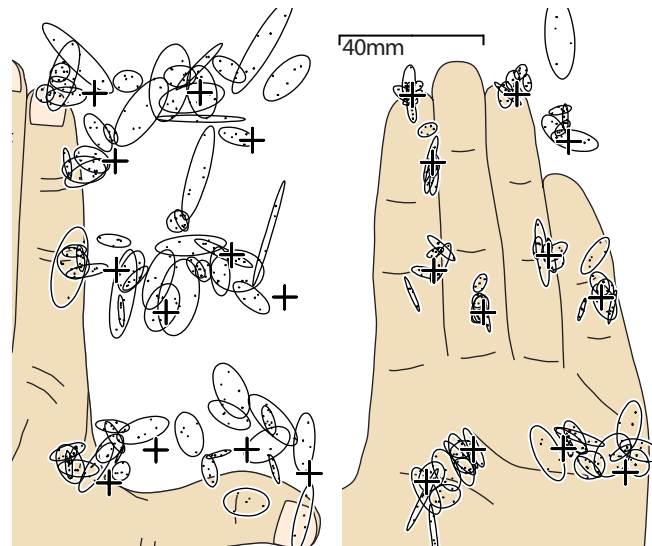

Figure 17: Study 2 results: all touches from all participants for (left) the *empty space* condition and (right) the *palm* condition. Plus signs indicate actual target positions. Ovals represent the bivariate normal distribution of selections per participant per target.

Figure 17 shows the raw data where each ellipse contains the 5 trials from one participant for one target. For the remaining analysis, we decomposed the targeting error (distance between target and acquired location) into sys-

tematic error (*offset*) and noise (*minimum button size*), as suggested by [8]. We assume one overall offset, instead of per user offsets. This is a more liberal estimate of touch accuracy, because it is not calibrated per user.

On average, the diameter of a circular minimum button on the *empty space* interface is 27.9mm (SD=0.32) and 17.7mm (SD=0.22) on the *palm* interface. This difference is statistically significant (t_{11}=2.912, *p*=0.014, Cohen's *d*=0.84).

When input is mapped back to the iPhone screen, all interactions are scaled down by a factor of 1.86 (the ratio of hand size to screen size). This in turn affects minimum button sizes—they now shrink to 15mm for the *empty space* condition and 9.5mm with the *palm* condition.

Discussion

As hypothesized, the *palm* interface was more accurate than the traditional *empty space* interface. This supports the reasoning presented earlier (see *Design Discussion*) and also confirms our choice to build the Imaginary Phone with a palm-based interface.

Also unsurprisingly (because we offered no feedback), the raw accuracy (i.e., before scaling down) obtained with the two imaginary interfaces is worse than the accuracy values obtained by other researchers with modern touchscreens. They report, using different study conditions, minimum button sizes of 15.0mm [8], 11.5mm [24] and 10.5mm [22]. The touch distributions we measured in this study indicate minimum button sizes of 9.5mm when scaled to fit the phone, which are comparable to a touch device. This, of course, is due to the palm offering a bigger input surface than the actual device, such that input errors shrink as input locations are mapped back to the device. In particular, this provides the necessary accuracy to acquire standard widgets on current touch devices, such as the iPhone, therefore making the transfer concept viable.

That said, these results were obtained with a tracking mechanism more accurate than what our current prototype can deliver. Consequently, these results could be considered a theoretical minimum for feedback-less targeting on the palm.

SUMMARY OF THE TWO STUDIES

Summarizing the two studies reported above, we find that all three of the assumptions of the transfer learning model stated earlier have support.

(1) Users indeed build up spatial memory automatically while using a physical device: participating iPhone users knew the correct location of 68% of their own iPhone home screen apps by heart.

(2) Spatial memory indeed transfers from physical to imaginary interfaces: participants recalled the location of home screen apps with 61% accuracy when pointing on the palm of their hand.

(3) Pointing on the palm is precise enough to allow operating the device: using accurate tracking, participants can reliably acquire targets less than 17.7mm in diameter on their palm. Mapping these back to the smaller iPhone increases precision to 9.5mm button sizes. This is sufficient to operate standard widgets on today's mobile touch devices.

We conclude that the transfer model is viable, even though full accuracy will not be redeemed until higher resolution tracking equipment becomes available.

CONCLUSIONS

In this paper, we presented a method of learning imaginary interfaces based on transfer and illustrated the concept with our Imaginary Phone prototype.

From the perspective of a mobile device user, the main benefit of imaginary interfaces based on transfer learning is that it saves users the effort for retrieving the device.
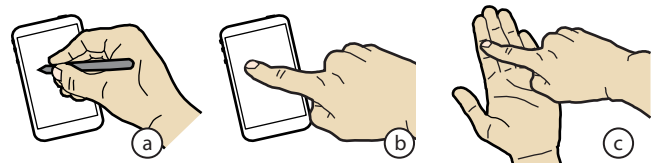

Figure 18: (a) Early mobile devices required users to retrieve a stylus and the device. (b) Current touch devices require retrieving only the device. (c) Imaginary interfaces do not require retrieving anything.

What is promising is that a similar transition has happened before, as illustrated by Figure 18. While early devices required users to retrieve device and stylus (e.g., PalmPilot), usage eventually transitioned to touchscreen-based devices. This move took place even though stylus input is in many ways superior to touch input—it offers higher precision (no fat finger problem [22, 8]).

At the expense of losing even more precision and essentially limiting interaction to microinteractions, systems like the Imaginary Phone have the potential to offer even more convenience, namely there is no need to retrieve the device anymore. While hardly viable on its own, we argue that the combination of an imaginary interface and a physical mobile device is an intriguing form factor.

As future work, transfer learning could be applied to a broader range of devices, such as remote controls and instrument panels. In particular, it would be interesting to investigate if the transfer learning principle can be applied to such devices (that have a strong tactile component) rather than the visual interface as we have shown here.

ACKNOWLEDGEMENTS

REFERENCES

1. Apple VoiceOver,
   http://www.apple.com/accessibility/iphone/vision.html

2. Ashbrook, D.L., Clawson, J.R., Lyons, K., Starner, T.E. and Patel, N. Quickdraw: the impact of mobility and on-body placement on device access time. In *Proc. CHI*, (2008), 219–222.

3. Ashbrook, D. *Enabling Mobile Microinteractions*. PhD Thesis. (2010).

4. Greiner, T.M. *Hand Anthropometry of U.S. Army Personnel*. U.S. Army Report ADA244533, (1991).

5. Gustafson, S. Bierwirth, D. and Baudisch, P. Imaginary Interfaces: spatial interaction with empty hands and without visual feedback. In *Proc. UIST*, (2010), 3–12.

6. Harrison, C., Tan, D. and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI*, (2010), 453–462.

7. Harrison, C. and Hudson, S.E. Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proc. UIST*, (2009), 121–124.

8. Holz, C. and Baudisch, P. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proc. CHI*, (2010), 581–590

9. Holz, C. and Wilson, A.D. Data Miming: inferring spatial object descriptions from human gesture. In *Proc. CHI*, (2011), 2501–2510

10. Kohli, L. and Whitton, M. The Haptic Hand: providing user interface feedback with the non-dominant hand in virtual environments. In *Proc. GI*, (2005), 1–8.

11. Kuester, F., Chen, M., Phair, M.E. and Mehring, C. Towards keyboard independent touch typing in VR. In *Proc. VRST*, (2005), 86–95.

12. Li, F.C.Y., Dearman, D. and Truong, K.N. Virtual Shelves: interactions with orientation aware devices. In *Proc. UIST*, (2009), 125–128.

13. Loclair, C., Gustafson, S. and Baudisch, P. PinchWatch: a wearable device for one-handed microinteractions. In *Proc. MobileHCI Workshop on Ensembles of On-Body Devices*, (2010), 4 pages.

14. Microsoft Kinect, http://www.xbox.com/kinect

15. Mistry, P., Maes, P., and Chang, L. WUW - wear Ur world: a wearable gestural interface. In *CHI Extended Abstracts*, (2009), 4111–4116.

16. Mistry, P. and Maes, P. Mouseless. In *Adjunct Proc. UIST*, (2010), 441–442.

17. Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proc. CHI*, (2005), 919–928

18. Perugini, S., Anderson, T.J. and Moroney, W.F. A study of out-of-turn interaction in menu-based, IVR, voicemail systems. In *Proc. CHI*, (2007), 961–970.

19. Rekimoto, J. GestureWrist and GesturePad: unobtrusive wearable interaction devices. In *Proc. ISWC*, (2001), 21–27.

20. Strachan, S., Murray-Smith, R. and O'Modhrain, S. BodySpace: inferring body pose for natural control of a music player. In *CHI Extended Abstracts*, (2007), 2001–2006.

21. Tamaki, E., Miyaki, T. and Rekimoto, J. Brainy Hand: an ear-worn hand gesture interaction device. In *CHI Extended Abstracts*, (2009), 4255–4260.

22. Vogel, D. and Baudisch, P. Shift: a technique for operating pen-based interfaces using touch. In *Proc. CHI*, (2007), 657–666.

23. Wachs, J.P., Kölsch, M., Stern, H. and Edan, Y. Vision-based hand-gesture applications. *CACM 54*, 2 (2011), 60–71.

24. Wang, F. and Ren, X. Empirical evaluation for finger input properties in multi-touch interaction. In *Proc. CHI*, (2009), 1063–1072.

25. Willis, K.D., Lin, J., Mitani, J. Igarashi, T. Spatial Sketch: bridging between movement & fabrication. In *Proc. TEI*, (2010), 5–12.

26. Yin, M. and Zhai, S. The benefits of augmenting telephone voice menu navigation with visual browsing and search. In *Proc. CHI*, (2006), 319–328.