Let's Kick It: How to Stop Wasting the Bottom Third of Your Large Scale Display

Ricardo Jota
Dept. of Computer Science
University of Toronto,
Toronto, Canada
jotacosta@dgp.toronto.edu

Pedro Lopes
Hasso Plattner Institute,
Potsdam, Germany
pedro.lopes@hpi.uni-potsdam.de

Daniel Wigdor
Dept. of Computer Science
University of Toronto,
Toronto, Canada
dwigdor@dgp.toronto.edu

Joaquim A. Jorge
VIMMI / Inesc-ID
IST
Lisbon, Portugal
jaj@inesc-id.pt

ABSTRACT
Large-scale touch surfaces have been widely studied in literature and adopted for public installations such as interactive billboards. However, current designs do not take into consideration that touching the interactive surface at different heights is not the same; for body-height displays, the bottom portion of the screen is within easier reach of the foot than the hand. We explore the design space of foot input on vertical surfaces and propose three distinct interaction modalities: hand, foot tapping, and foot gesturing. Our design exploration pays particular attention to areas of the touch surface that were previously overlooked: out of hand's reach and close to the floor. We instantiate our design space with a working prototype of an interactive surface, in which we are able to distinguish between finger and foot tapping and extend the input area beyond the bottom of the display to support foot gestures.

Author Keywords
Large-scale display, foot interaction, kick, floor input.

ACM Classification Keywords
H.5.2. [Information interfaces and presentation]: User Interfaces. – Graphical user interfaces.

INTRODUCTION
Researchers foresee a future in which all walls, windows, doors – indeed, all vertical surfaces – hold the potential to serve as interactive displays. When such a future arrives, these displays will have uses far beyond the current commercial niches of telepresence [9] and advertising [7]. Several projects have explored interactions on such large-scale interactive surfaces. One area of focus has been on reaching far-away targets [16]. While useful, these techniques are primarily addressing horizontal distance: reducing the time spent moving an object from a horizontal location to another. Largely unexplored is the issue of *vertical* position – that is, how an object's position on the large-scale display affects the types of interactions employed to use the display.

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

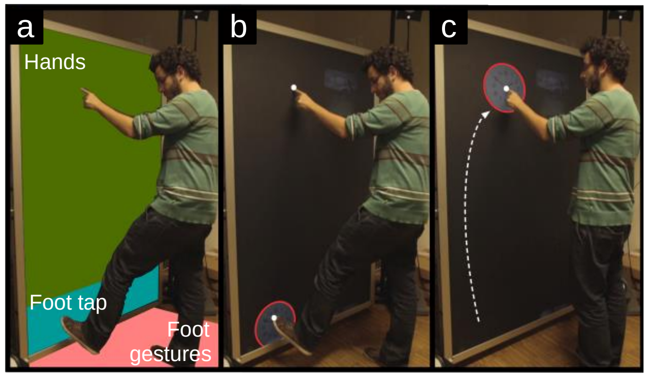

Figure 1: (a) We propose three types of input, each ergonomically fitting one interactive region: hand interaction, foot taps, and foot gestures; (b-c) shows an example technique: kicking an object in the lower portion of the display causes it to pop-up to finger position.

In the present work, we address exactly this issue: "how to facilitate interaction with the lower region of touch displays, which is beyond hand-reach?". To explore this concern, we developed a vertical touch surface, sensitive to both feet and finger input, which is depicted in Figure 1. In our early explorations, it became immediately apparent that, for adult users, the area towards the bottom of the screen was far more easily reached with the feet than with the hands; we often found ourselves kicking at items on the display to avoid bending down to touch, as shown in Figure 1(b). This is aligned with ergonomic literature, which regards bending down as a cause for back strain; and therefore, to be avoided [3]. What was also evident, though, was that traditional direct-physical manipulation techniques were not suitable for foot-based interaction. Thus, in this work, we propose an adaption of direct manipulation techniques that takes into account the ergonomics of foot input.

As current touch-surfaces are not able to distinguish hand from feet contacts, we expanded the sensing capabilities of our prototype by adding a contact microphone, which enables us to acoustically differentiate tapping with the foot or with the hand or fingers. Furthermore, a depth-sensing camera allows us to detect foot position and orientation beneath and around the bottom of the display. Using these input streams, we developed three types of foot-based interaction methods: (1) foot tapping; (2) foot gesturing; and, (3) multi-modal foot+hand combinations (as shown in Figure 1).

We condense our input techniques by proposing a design space for direct touch input on large-scale displays, taking input appendage and ergonomics into account. We review prior work with foot input, explore six interaction methods making use of foot input, and describe the prototype technical details. Finally, we discuss findings and observations obtained during the exploration.

DESIGNING FOR FOOT INPUT

Inspired by research in ergonomics [17], we identified three distinct interactive regions: hand, foot tapping and foot gesturing, depicted in Figure 1(a). These regions are not used for differentiating input; our sensors do that. Instead, they guide our interaction designs, by taking into account the ergonomics of inputs closer to those areas.

Hand: This region extends from the top of the display to below the user's waist, where a user close to the display can still reach down comfortably. Within this region, traditional direct touch physical manipulation can be used, along with reaching techniques described elsewhere.

Foot Tapping*:* This region is below the knee level, which would otherwise require kneeling or crouching to operate with the hands. We note that humans are less dexterous with their feet than with their hands [4]. Our findings indicate that the rich set of gestures performed by the hands should be reduced to a small set of interactions accomplished by only *tapping* with the foot (i.e., no sliding/flicking, and so forth). It is also appropriate to limit the precision of foot-tapping gestures. Although foot tapping may occur on the screen, where finger tapping can be pixel-precise, foot gestures are only used to specify whole objects, and not specific points within those objects, as is the case with the equivalent finger gestures.

Foot Gesturing*:* In our prototype, we intentionally left a 10 cm gap between the floor and the display. We found that this space allows an increase in the expressiveness of input to this region since users can make gestures in front of the display, using techniques such as sliding the foot under it or pulling the foot away from it. This is possible because input to this area is made while both feet remain supported, unlike the tapping, which requires a foot to be lifted. Since this region is below the display (and not actually part of it), the key challenge regards with how to extend the existing "direct" manipulation metaphors beyond the screen.

RELATED WORK

Given the promise of large displays for interaction, significant efforts have been expended addressing the issue of acquiring out of reach targets on large displays. These include using indirect pointing methods and providing tools to 'pull' distant targets closer [2, 16]. While these all serve to reduce the physical distance from users to targets, their use for targets in front of but below the reach of a user would mean failing to take advantage of the pleasing directness of foot-based interaction.

In developing our design space, we built on two areas of earlier work: foot input, and contact type differentiation.

Foot Input

Augsten et al. shows how high precision foot input can be made available through FTIR-illuminated floors [1] and demonstrates how to locate the position of a foot held in mid air above the floor. Han et al. add kicking to mobile interactions by using the device camera to detect kicks by observing leg motion from above [5]. Finally, Schoening et al. combine finger input with foot poses, using a balance board, to manipulate data on a vertical surface [8]. Our project builds on these studies by using the foot for differentiated direct-touch manipulation of the full vertical screen thereby enabling the large display surface from top to bottom.

Appendage Differentiation

We distinguish foot from hand input using acoustic sensing. Similarly, researchers have used acoustic cues to distinguish between different finger regions, but never implement it for feet vs. finger identification. Paradiso et al. used the frequency signature of the touch contacts to distinguish between knuckles and fingertips [13]. These acoustic cues have been used on touch surfaces to distinguish between finger pads, fingertips, knuckles, and fingernails [6,11]. Also, Touché uses capacitive sensing with sweeping frequency to distinguish the electrical signatures of different poses; however, the technique is limited to direct skin contacts [14]. Additionally, shape has been used to infer input style, such as in Wu et al. shape sensing for multi-touch surfaces [1]. We build upon these projects by distinguishing between contacts made with the screen using a feet or fingers through acoustic sensing.

INTERACTION TECHNIQUES

We propose a set of six interaction techniques, iteratively designed and implemented on our prototype, empowering users with access to the unutilized lower portion of a large-screen display using direct-touch gestures with their feet.

Kicking & Throwing

It is common in windowing systems to provide a method to temporarily dismiss (or 'minimize') an application. After being dismissed, these applications are then iconified in the Task Bar (Windows 95-8), or in the Dock (OSX). Similarly, we provide a mechanism to quickly transition an object between the Hand region and the Foot Tapping region

1. Throw to the Bottom*:* Users "throw" objects down by flicking them with handed input towards the bottom of the display, this causes the object to transition to the feet-level region. This is similar to the gesture demonstrated by Wu and Balakrishnan to pass objects across an interactive tabletop [1]. This technique provides a fast way to lower objects using a direct physical manipulation metaphor and, at the same time, does not require the user to lean down.

***2. Kick to Eye-Level*:** To interact with an object below arm reach, a user simply kicks the object. This causes the object to be moved upwards, near eye-level. The object may be returned to the bottom of the screen using the throwing gesture, or the foot+hand method described below.

***3. Relocate with Foot+Hand**:* Greater expressivity is achieved by utilizing both the foot and the hand. In the previous two gestures, the horizontal location of the object was invariant: users could only change the vertical position, and even then only to a fixed value. To allow greater control, users can simultaneously touch an object and an empty space for relocation. The object and destination can be specified with either the hand or foot, allowing for greater position when throwing to the bottom of the screen, or when kicking it upward as illustrated in Figure 1.

Extending to the Floor

The 10cm gap beneath the display offers the user the area below the screen for foot input. A depth-sensing camera is utilized to enable the following gestures.

4. Object Glimpse: The user can 'lift' an object with their toes to temporarily move objects upwards to facilitate further finger gestures. The user slides a foot under the display and lifts her toes upwards, with the heel anchored, as shown in Figure 3. This action slightly raises all objects directly above the foot for the duration of the gesture. When the user removes the foot, objects slowly animate back to their original position, allowing the user an opportunity to quickly move them to eye level.

5. Rearrange Thrown Objects: Users affect the horizontal movement of objects on the bottom of the screen using a foot swipe. For such, users place their foot beneath the object on the screen and then slide it horizontally to the desired location (as illustrated in Figure 3).

6. Delete Objects: To delete an object from the display, we utilized a metaphor of stepping on the image and dragging it out. To achieve this, users slide their foot under the display and then drag it backwards.

IMPLEMENTATION

We built our prototype device, as depicted in Figure 4. It is 1.7m tall by 1m wide. As such, it is significantly narrower than traditional large-scale displays. Nonetheless, its size is sufficient to allow us to explore interaction methods. Multitouch input is sensed using a laser-light plane (LLP) working on the infrared spectrum and captured via a camera. A microphone mounted at one of the four corners, depicted in Figure 5(a), allows us to capture the sound of surface contacts; these are then used to classify finger or foot touches by analyzing the frequency signature of each contact, as described in [6,11]. This works because the frequency distribution of the kick sound is inherently different from a finger tap. The microphone, in Figure 4(a), is sampled in 64 ms chunks; then, the sample's 11-band FFT is compared against the database of pre-trained gestures using the algorithm described in [11]. To train the database we collected 10 x 2 (both hands) finger taps (scattered around the whole surface) and 10 x 2 (both feet) foot taps against the lower region. During informal design sessions, the recognition rate was always above 90%.

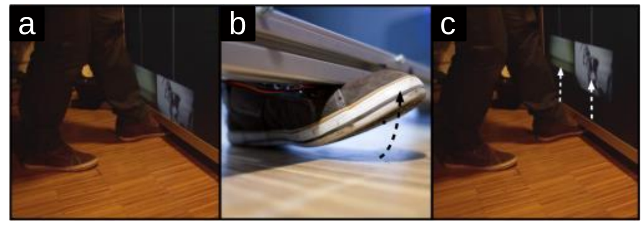

Figure 2: Toe lift gesture: Users slide the foot under the display (a), lifts his toes (b), and objects rise (c).

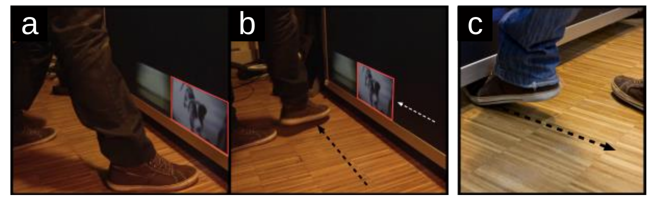

Figure 3: Foot slide: the user places their foot under the screen (a) and slides it horizontally. The object follows (b). Foot Pull: the user removes objects by pulling them away from screen.

Finally, a depth-camera is mounted at the back of the frame to detect foot input beneath and in front of the display, it provides an un-occluded view of the user's foot when gesturing. Tracking the feet is realized using a custom 3D blob tracker using real-time point cloud data from the depth camera, which reports blobs (corresponding to feet) that are within 20 cm of the display. The tracking is as follows: all data that does not reside within the defined interaction space is filtered; then, point cloud data is clustered into blobs that must exceed a radial threshold size (empirically optimized to 4 cm) to be classified as a foot, if multiple blobs are found, the closest to the display is used. Lastly, using the *mt4j* multitouch library, foot gestures (e.g., swipe, toe lift) are recognized by their unique 3D spatial pattern.

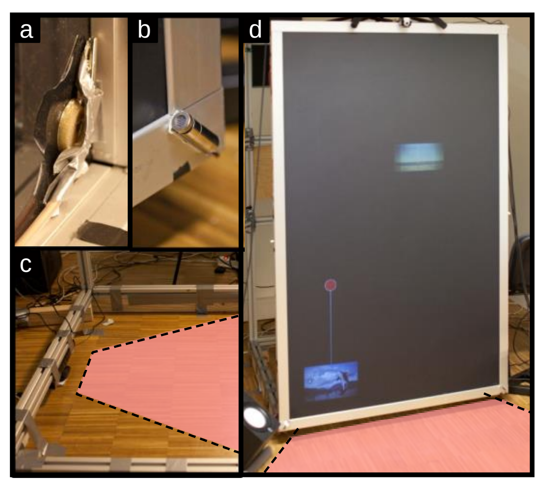

Figure 4: (d) The prototype is comprised of: (a) surface microphones for classifying finger vs. foot input.;(b) IR light laser plane generators for optical multitouch; and (c) a depth camera looking under the surface to capture feet gestures.

FOOT+HAND DISCUSSION

Techniques made possible with foot input have not been exhausted. Future contributions to this design space need to consider the following: it is clear that several input primitives are less suitable for fancy foot work, than for direct-touch hand input. For example, dragging a finger across the display is a common input action; however, dragging a foot it is difficult, since the user is required to balance on the other foot. We found out that finger dragging cannot be directly mapped to its foot-equivalent, thus should be re-designed (e.g., using multiple kicks).

Further, tap-based interactions need to be adjusted for foot-based input. We found, unsurprisingly, that users were less accurate when tapping with their feet than with their hands. Thus, we recommend expanding the 'iceberg' [19] targeting areas for foot interaction beyond those used for hand input. In target-dense environments, it may be necessary to limit the number of targets available for foot input, due to overlap of targeting areas.

While our interaction techniques extend direct touch input by mapping foot input on the bezel onto the adjacent pixels, alternative form factors might otherwise extend direct-touch. For example, previous projects have demonstrated the potential for non-flat displays to enhance desktop experiences [18]. A curved wall display could be integrated with a floor display to unify tapping and floor gestures into a continuous space, changing the interaction vocabulary. While such a solution would occlude the mount point for our depth camera, floor based sensors [1] and see-through displays could allow imaging of foot positions [20].

CONCLUSIONS

We have demonstrated foot input for the bottom portion of large-scale displays. We described several interaction techniques, which improve existing direct-touch interaction with the bottom of the display. Lastly, we proposed several critical considerations to be made when designing interactive systems that take advantage of foot input.

ACKNOWLEDGMENTS

We thank Kim Wright and Katie Barker for their revisions. Partially funded by FCT CEDAR PTDC/EIA/116070/2009.

REFERENCES

1. Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C., and Baudisch, P. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. *UIST '10*, 209-218.
2. Banerjee, A., Burstyn, J., Girouard, A., and Vertegaal, R. Pointable: an in-air pointing technique to manipulate out-of-reach targets on tabletops. *ITS '11*, 11-20.
3. Bridger, R., Some fundamental aspects of posture related to ergonomics, *Industrial Ergonomics '91*, 3–15.
4. Chan, K., Chan, A. Spatial stimulus–response compatibility for hand and foot controls with vertical plane visual signals. *Displays, 32:5, 2011*, 237-243.
5. Han, T., Alexander, J., Karnik, A., Irani, P., and Subramanian, S. Kick: investigating the use of kick gestures for mobile interactions. *MobileHCI '11*, 29-32.
6. Harrison, C., Schwarz, J., Hudson, S. TapSense: enhancing finger interaction on touch surfaces. *UIST '11*, 627-636.
7. Jacucci, G., Morrison, A., Richard, G., Kleimola, J., Peltonen, P., Parisi, L., and Laitinen, T., Worlds of information: designing for engagement at a public multi-touch display. *CHI '10*, 2267-2276.
8. Schoening, J., Daiber, F., Krueger, A., and Rohs, M.. Using hands and feet to navigate and manipulate spatial data. Proc., *CHI EA '09*, 4663–4668.
9. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., and Vertegaal, R. TeleHuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. *CHI '12*, 2531-2540.
10. Koppel, M., Bailly, G., Müller, J., and Walter, R. Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. *CHI '12*, 317-326.
11. Lopes, P., Jota, R., Jorge, J. Augmenting touch interaction through acoustic sensing. *ITS '11*, 53-56.
12. Müller, J., Walter, R., Bailly, G., Nischt, M., and Alt, F.. Looking glass: a field study on noticing interactivity of a shop window. *CHI '12*, 297-306.
13. Paradiso, J., Leo, C., Checka, N., and Hsiao, K. Passive acoustic sensing for tracking knocks atop large interactive displays. *IEEE ICS*, 11–14.
14. Sato, M., Poupyrev, I., and Harrison, C. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. *CHI '12, 483-492.*
15. Scott, J., Dearman, D., Yatani, K., and Truong K. Sensing foot gestures from the pocket. *UIST '10*. 199-208.
16. Schick, A., Camp, F., Ijsselmuiden, J., and Stiefelhagen, R. Extending touch: towards interaction with large-scale surfaces. *ITS '09*, 117-124.
17. Toney, A., and Thomas, B. Considering Reach in Tangible and Table Top Design. *TABLETOP '06*, 57-58.
18. Weiss, M., Voelker, S., Sutter, C.,and Borchers, J. The BendDesk: dragging across the curve. ITS '10, 317-317.
19. Wigdor, D., and Wixon, D. Brave NUI World (1st ed.). Morgan Kaufmann.
20. Wilson, A. TouchLight: an imaging touch screen and display for gesture-based interaction. ICMI '04, 69-76.
21. Wobbrock, J., Morris, M., and Wilson, A. User-defined gestures for surface computing. *CHI '09*, 1083-1092.
22. Wu, M., and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. *UIST '03* 193-202.