

Spatial Gestures using a Tactile-Proprioceptive Display

Eelke Folmer and Tony Morelli
Player-Game Interaction Lab
University of Nevada, Reno
{efolmer, morelli}@cse.unr.edu

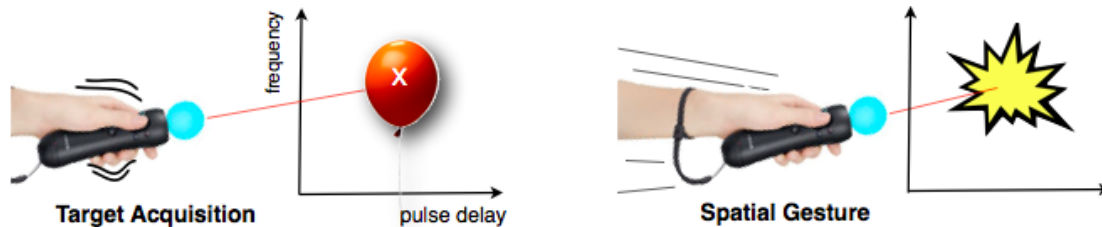


Figure 1. (Left) The user scans a plane that is defined in front of them to locate the X and Y of a target that is indicated using directional vibrotactile feedback. (Right) Because the controller points towards the target, the user can perform a spatial gesture towards the target, using proprioception.

ABSTRACT

Proprioception—the human ability to sense the orientation of limbs without vision or hearing—is one of the main drivers of motor operations and plays a significant role in input modalities such as touch and gestures. As an output modality proprioception has remained largely unexplored—though it can convey information to a user using their own body. Spatial interaction requires users to visually acquire the location of an object, which can then be manipulated using a touch or gesture. This is challenging if you are unable to see or in mobile contexts where the use of a display may be undesirable. This paper evaluates a tactile-proprioceptive display for eye and ear free 2D target acquisition and spatial interaction.

ACM Classification Keywords

H.5.2 [User Interfaces]: - Haptic I/O

General Terms

Design, Human Factors, Experimentation

Author Keywords

Non-visual NUI, Spatial Gestures, Haptics, Accessibility

INTRODUCTION

Spatial interaction is a key component of natural user interfaces (NUI), where a touch or a gesture activates or alters the position of on-screen objects. Spatial interaction relies upon being able to visually acquire the position of an object, which is challenging for a user who is visually impaired [6] or in particular mobile contexts where the use of a display may be dangerous or inappropriate [2]. In a quest for maximum mobility recent research efforts seek to remove screens

altogether [3, 5]. Several non-visual NUI's have been proposed [2, 4] yet these only facilitate non-spatial gestures [3], such as navigating a list and selecting an item. Recently spatial interaction has been brought to screen-less interfaces [3, 5]. These interfaces require the user to keep track of object positions using their visuospatial memory. The lack of visual spatial feedback may impose a significant cognitive load on the user—especially when multiple objects are present and visuospatial memory fades over time. It also doesn't allow for users to recover the location of an object if they are unable to remember it. Though audio [3] coordinates have been explored for spatial feedback this may be difficult to employ in mobile contexts due to noise, privacy or social concerns.

This paper addresses the current limitations of existing non-visual NUI's by presenting an eye and ear free *display* technique that can point out the location of an object in a 2D display defined in front of the user, and which allows for the user to manipulate this object using a spatial gesture. Our display uses *proprioception*—the human ability to sense the position and orientation of their limbs—to appropriate the human body into a display device. Haptic feedback can be augmented with proprioceptive information to facilitate a significantly larger information space that can be accessed in an ear and eye free manner. For example, *tactile-proprioceptive displays* [1, 6] have been explored to point out a target. Users scan their environment with a mobile device and a vibrotactile cue guides the user to point their device at the target. Target direction is then conveyed to the user using their own arm; effectively appropriating the human body into a display.

Prior research on tactile-proprioceptive displays has only explored 1D target acquisition. We first evaluate two scanning strategies for 2D target acquisition and then evaluate the users' ability to perform a spatial gesture at the target. Our NUI could have useful mobile-, assistive technology- or rehabilitation -applications, for example, to facilitate exercise games for users who are blind [6] or develop low-cost approaches towards motor rehabilitation or sports training.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

TEI 2012, February 19 - 22, 2012, Kingston, Ontario, Canada.

Copyright 2012 ACM 978-1-4503-0541-9/11/08-09...\$10.00.

RELATED WORK

Research in developing non-visual NUI's has initially focused on exploring how touch screens can be made accessible to users who are blind, for example, through the definition of custom gestures [4]. To achieve mobility, several screen-less mobile interfaces have been developed, for example, the Nanya ring [2] uses physical motion for input. Due to the lack of spatial feedback these approaches only allow for non-spatial gestures, e.g., gestures that manipulate high level container-like widgets, such as pages or lists, but not individual objects. A different approach exploits users' visuospatial memory. Virtual shelves [5] is a non-visual input technique where users trigger shortcuts from spatial memory by positioning a Wii remote within a circular hemisphere in front of them. Virtual shelves does not perform spatial gestures on individual objects. Gustafson presents an imaginary interface [3] where virtual objects defined in a plane in front of the user can be manipulated using spatial gestures. Users are required to keep track of each object's position, which may be challenging to perform especially when multiple objects are present. Though Gustafson explores an audio based coordinate system to indicate an object's location, audio may be difficult to use in mobile contexts, which are often subject to noise, privacy or social concerns.

Tactile feedback lends itself well to achieve eye and ear free interaction [8] as mobile devices are often held closely to our skin. A number of tactile-proprioceptive displays have been explored for 1D target acquisition. Sweep-Shake [9] is a mobile application that points out geolocated information. The direction the phone is pointing is determined using a compass and GPS. Directional vibrotactile feedback (pulse delay) is used to render a target. A study with four users found users could locate a target on a 360° horizontal circle in 16.5 seconds. Ahmaniemi [1] explored target acquisition using a high precision inertial tracker (gyroscope, compass and accelerometer) and a C2 vibrotactor. Directional and non-directional vibrotactile feedback (frequency and amplitude) were explored for rendering targets with varying sizes on a 90° horizontal line. A user study with eight sighted users found they were able to find targets on average in 1.8 seconds. Target sizes larger than 15° were found to be most effective. Directional feedback was found to be more efficient than non-directional when target distance is furthest, but it negatively affects finding targets that are close. In prior work we combined gesture based interaction with 1D target acquisition using a proprioceptive display. VI Bowling [6] is an exercise game for users who are blind. A tactile-proprioceptive display was facilitated using a Wii remote where directional vibrotactile feedback (pulse delay) directs the player to point their controller at the location of the pins. Once the location of the pins is acquired, users can perform a gesture that resembles throwing a bowling ball as to hit the pins. With a close-to-target window of 38.6° and a target size of 7.2° a user study with six legally blind adults found that a target could be found on average in 8.8 seconds and users could perform a gesture with an average aiming error of 9.8°.

STUDY 1: 2D TARGET ACQUISITION

Instrumentation: Previous work on tactile-proprioceptive displays [1] used a high-precision inertial tracker, which works well for pointing but these are subject to drift over time [10] –which is the case when spatial gestures are performed with them. Instead we implemented a tactile-proprioceptive display using a commercially available game controller (Sony Move) whose location can be tracked using a camera. The controller's orientation is tracked using inertial sensing and it features an LED that serves as an active marker. The uniform spherical shape and known size of the LED allows for controller's position to be tracked in three dimensions with high precision (1 millimeter) using an external camera, i.e, the PlayStation Eye, which captures video at 640x480 (60Hz). Its integrated vibrotactor can provide directional vibrotactile feedback using pulse delay and frequency (91 to 275Hz). A 2D display is defined in front of the user whose width is determined by the reach of the players' arm, which gives it an effective width of 180° and a height of 135°. Two different scanning strategies were defined:

In *multilinear* scanning, directional vibrotactile feedback is provided on both axes simultaneously to indicate the target. Frequency is used for the Y and pulse delay for X. A related study with 1D target selection using a haptic mouse [7] found that targets can be found significantly faster when the difference between the on-the-target cue and close-to-target cue is increased at the border of a target. We varied the frequency linearly based on the Y target error with a maximum value of 140Hz at the edge of the target and 275Hz when on target. The pulse delay was 0ms when on target and 200ms at the edge, which decreased linearly with 3 ms/pixel. These values were established through play testing. Ahmaniemi [1] found target sizes larger than 15° to be most effective for a range of 90°. The Move controller is tracked in 640x480 resolution, therefore a target size of 100x80 pixels is used to have a similar target size.

Initial trials found multilinear scanning challenging to perform therefore a simpler form of scanning was defined. In *linear scanning* the user acquires the X and Y coordinate of the target sequentially. A horizontal band of 100 pixels is defined around the target's X coordinate where vibrotactile feedback is provided and frequency modulation indicates the target's Y coordinate. A space invaders like game was created in which players must find a random target and shoot it by pulling the trigger on the controller. The use of a game was motivated by that they are generally considered to be powerful motivators, which may allow for measuring optimal performance. The game runs on a laptop and communicates with the Playstation 3 to retrieve the controller's location and to adjust the haptic feedback. Targets are destroyed if the player fires a shot within 150 pixels of the target. The faster a target is destroyed the more points a user scores. A target was defined randomly in the display excluding a 20 pixel border to avoid scanning near the edge of the display. If the user moved the controller outside of the display the haptic feedback would stop. The use of random targets as opposed to fixed targets was motivated by that it allows us state the general performance of a proprioceptive display, in-

dependent of a specific application. When a target was hit, a sound effect would be played, haptic feedback would stop and score would be announced using speech. Background music was played to mask the sound of the vibrotactor.

Participants: We recruited 16 computer science students (1 female, average age 28.9, $SD=6.2$). Participants were randomly assigned into two eight-people groups (A,B) where group A played the game using linear and group B using multilinear scanning. All subjects were right handed and none reported any non-correctable impairments in perception or motor control. We measured players' height ($M=175.3$ cm, $SD=4.8$) and arm length ($M=59.2$ cm $SD=4.9$).

Procedure: Participants played the game using their dominant arm while standing. An observer was present during the study. Due to players having different heights and different arm lengths, we calibrated the position of the player, such that the horizontal range of their arm matched the horizontal range of the camera using a visual task. After calibration, we avoided players moving by requiring them to stand on a piece of paper. Players were then instructed what the goal of the game was and how to play the game either using linear or multilinear scanning. Players were allowed to briefly familiarize themselves with their scanning technique. While playing the game only haptic and audio feedback was provided. Players then played the game until 40 targets were hit. The game logged: (1) all target locations; (2) locations and orientations of the controller (sampled with 60 fps) including time stamps; (3) all shots fired and distance from the target when the shot was fired. Participants received \$5 gift for their participation.

Results: We consider the first 10 targets to be part of the learning phase, therefore we restrict our analysis on the players' performance of acquiring the last 30 targets. The average search time for a target was 9,214 ($SD= 2,691$) ms for linear and 7,712 ms ($SD=2,827$) for multilinear scanning. This is a relative unfair comparison as targets were generated at random and the target distance may vary significantly between both techniques due to randomness. Search time corrected for distance was: 51.7 ms/pixel ($SD=12.9$) for linear and 40.3 ms/pixel ($SD=12.0$) for multilinear scanning. This difference was found to be significant ($T_{2,14} = -2.298$ $p = .038$). Error was defined as the number of shots fired by the player that did not hit the target. Due to an error one dataset of errors for one user for multilinear scanning was lost. Linear scanning yielded an average error of 1.0 ($SD=0.3$) per target and multilinear scanning found an error of 1.5 ($SD=1.0$) per target. These values were not found to be significantly different ($T_{2,13} = .749$ $p > .05$). For multilinear 54% ($SD=15$) of the targets and 63% ($SD=17$) of the targets for linear were hit on the first attempt, which reveals that if the player's first shot missed, it often took multiple subsequent attempts to hit the target. A possible explanation is that players would randomly fire shots in the area where they first found the target rather than reacquiring the target. If a player successfully hit the target the controller was on average within 72.8 pixels for linear and 75.3 pixels for multilinear scanning of the target.

STUDY 2: PERFORMING SPATIAL GESTURES

Based on the results of the first study, we explore performing a spatial gesture towards an acquired target.

Instrumentation: Whereas in the first study a target is defined in the camera's frame (exocentric) in this study we extend the target's location with an egocentric (relative to the user) component. The target's X and Y are defined in the display and its Z is defined by the orientation of the controller. This setup allows for evaluating the accuracy with which a user can perform a spatial gesture aimed at a target that is conveyed using proprioception. A second game was created in which players use multilinear scanning to find a target (balloon). Instead of shooting it, players make a thrust gesture in the direction the controller is pointing as to pop the target. The accuracy of the gesture is measured by camera and the controller's gyroscope. The orientation (ω) of the controller is reported in Euler angles (θ, ϕ, ψ) where θ is yaw, ϕ is pitch and ψ is roll (see Figure 2). We used the same target size as in the first study. The position of the sphere is reported in pixels for the X and Y coordinate and in millimeters for its Z coordinate. The gesture was considered successful if: (1) it started within 150 pixels of the target; and (2) the Z-axis position of the controller decreased by 200 mm; and (3) the orientation of the controller remained within 5% error on each of the Euler angles throughout the gesture. These values were determined through play testing. Sound effects were used to indicate a hit or a miss. The faster targets are hit the higher the player's score.

Participants: We recruited 8 users (1 female, average age 28.13, $SD=3.56$) to participate in a user study. None of these participants participated in the first user study. All subjects were right handed and none reported any non-correctable impairments in perception or motor control. Players' height ($M= 179.3$ cm, $SD=9.5$) and arm length ($M=60.5$ cm $SD=6.2$) were measured.

Procedure: Where in the first study players held the controller with their arm stretched, for this study we required them to scan the plane using their forearm as to allow for making a thrust gesture. We calibrated the position of the player so that the horizontal and vertical range of their forearm matched the entire range of the camera. Players were instructed to hold the controller in the anterior position, i.e., with the palm facing up and to thrust the controller in the direction it was pointing once they found the target. Players were taught how to search for the target using multilinear scanning. Players could familiarize themselves with the technique and pop a few targets until they felt comfortable performing the procedure. Players played the game until they successfully popped 40 balloons. Participants received \$5 gift for their participation.

Results: Similar to study 1 we focus our analysis on the last 30 targets. The average time to perform a spatial gesture that hits the target was 10,322 ($SD=2,542$) ms. The average number of misses per target was 1.6 ($SD=0.7$) and 63% ($SD=14$) of the targets were hit on the first attempt.

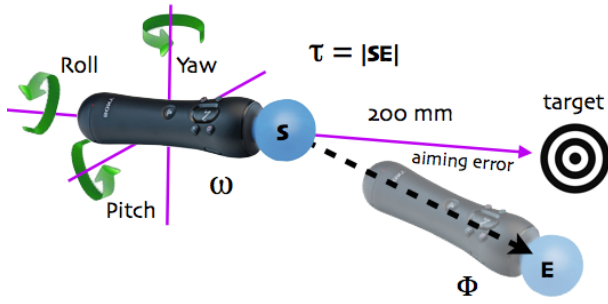


Figure 2. Orientation of the controller using Yaw, Pitch and Roll and computation of the real gesture error due to drift

4.1% of misses were due to too large deviations in the controller's orientation with one user responsible for 87.5% of this type of miss. The remaining 95.9% of the misses were due to not being on target prior to making the gesture. When hit the controller's position was within 74.3 pixels of the target prior to making the gesture. Corrected search time was 45.9 (SD=9.9) ms/pixel. We found the following average normalized change between the start (ω) and end (Φ) orientation of each successful gesture: θ : 3.2° (SD=3.1), ϕ : 2.9° (SD=2.7), ψ : 3.6° (SD=3.8). This corresponds to an axis angle rotation, i.e., $\omega \cdot \Phi$ of 5.6° . No significant difference ($P > .001$) was found between the average Euler angles, which indicates that an error in orientation is just as likely on any of the Euler angles. Gestures may be subject to drift, therefore to compute its real error, we compute the vector that represents the displacement of the controller during the gesture: $\tau = |\vec{SE}|$ where S is the start and E the end position (see Figure 2). We converted the X and Y coordinates from pixels to millimeters using the user's arm length. We then compute $\omega \cdot \tau$ (with ω the initial orientation) to find an average gesture error of: 21.4° (SD=13.2).

DISCUSSION AND FUTURE WORK

The average search time for multilinear scanning was four times as large as what Ahmaniemi [1] found, but it was comparable with our previous results [6] –though our display was two dimensional and significantly larger in size. In prior work we found an aiming accuracy of 9.8° but this is a different type of gesture (throw versus thrust) whose error was measured in 1D, which makes comparison with the 3D error of 21.4° in this study difficult. Multilinear scanning was found to be significantly faster than linear scanning. In retrospect this may seemed trivial to investigate as users are presented with more information, though based on preliminary trials legitimate concerns were raised about the learnability of multilinear scanning, which the user studies were able to contradict. Some players spent time searching for the target outside the range of the camera or failed to scan all the way to the edge of the display, which was detrimental to their performance. Rather than interrupting haptic feedback, we need to more clearly indicate to the user when the controller is at the edge of the display.

A limitation of using a controller for spatial interaction is that we can only do simple gestures, such as taps and not more sophisticated hand gestures, such as pinches or pan. Future work will explore using a smaller controller attached

to the user's wrist leaving the hand free for providing gestures. Imaginary interfaces [3] uses one hand to indicate a point of reference to the camera, limiting spatial interaction to a single hand. Because we use active markers our approach can easily be extended to facilitate spatial gestures using both hands. Future work will also explore 3D target selection using a controller capable of providing a third type of directional vibrotactile feedback (amplitude) or using a second controller to convey the target's Z-coordinate. Of interest to investigate would be if users are cognitively capable of scanning to two 2D targets simultaneously and if they can scan a 3D target along three axes simultaneously.

ACKNOWLEDGEMENTS

This work is supported by NSF Grant IIS-1118074.

REFERENCES

1. T. T. Ahmaniemi and V. T. Lantz. Augmented reality target finding based on tactile cues. In *Proc. of ICMI-MLMI '09*, pages 335–342, New York, NY, USA, 2009.
2. D. Ashbrook, P. Baudisch, and S. White. Nanya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proc. of CHI '11*, pages 2043–2046, New York, NY, USA, 2011. ACM.
3. S. Gustafson, D. Bierwirth, and P. Baudisch. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proc. of UIST '10*, pages 3–12, New York, NY, USA, 2010. ACM.
4. S. K. Kane, J. P. Bigham, and J. O. Wobbrock. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings ASSETS'08*, pages 73–80, New York, NY, USA, 2008. ACM.
5. F. C. Y. Li, D. Dearman, and K. N. Truong. Virtual shelves: interactions with orientation aware devices. In *Proceedings of UIST'09*, pages 125–128, New York, NY, USA, 2009. ACM.
6. T. Morelli, J. Foley, and E. Folmer. VI-Bowling: a tactile spatial exergame for individuals with visual impairments. In *Proc. of ASSETS'10*, pages 179–186, New York, NY, USA, 2010. ACM.
7. T. Oron-Gilad, J. L. Downs, R. D. Gilson, and P. A. Hancock. Vibrotactile guidance cues for target acquisition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 37(5) pages 993–1004, 2007.
8. I. Poupyrev, S. Maruyama, and J. Rekimoto. Ambient touch: designing tactile interfaces for handheld devices. In *UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 51–60, New York, NY, USA, 2002. ACM.
9. S. Robinson, P. Eslambolchilar, and M. Jones. Sweep-shake: finding digital resources in physical environments. In *Proc. of MobileHCI'09*, pages 12:1–12:10, New York, NY, USA, 2009. ACM.
10. D. Vlastic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popović. Practical motion capture in everyday surroundings. In *proc. of SIGGRAPH '07*, page 35, New York, NY, USA, 2007. ACM.