Haptic Target Acquisition to Enable Spatial Gestures in Nonvisual Displays

Alexander Fiannaca* Tony Morelli† Eelke Folmer‡

Department of Computer Science and Engineering - University of Nevada, Reno, USA

ABSTRACT

Nonvisual natural user interfaces can facilitate gesture-based interaction without having to rely on a physical display. Consequently, this may significantly increase available interaction space on mobile devices, where screen real estate is limited. Interacting with invisible objects is challenging though, as such techniques do not provide any spatial feedback but rely entirely on users' visuospatial memory. This paper presents an interaction technique that appropriates a user's arm using haptic feedback to point out the location of nonvisual objects; thereby, allowing for spatial interaction with nonvisual objects. User studies evaluate the effectiveness of two different single-arm target-scanning strategies for selecting an object in 3D and two bimanual target-scanning strategies for selecting an object in 2D. Potential useful applications of our techniques are outlined.

Keywords: Pointing, Gestural input, Natural language interfaces, Haptic devices.

Index Terms: H.5.2 [Haptic Interfaces]—Haptic I/O;

1 INTRODUCTION

Natural user interfaces (NUIs) seek to capitalize on the innate abilities that users have acquired through interactions with the real world by removing intermediary devices, such as a keyboard and mouse, so as to facilitate an invisible interface that is more intuitive to use. NUIs predominantly define novel input modalities [30], such as touch, gestures, and speech, but recent work [8, 17, 23] has explored gesture-based interaction without using a visual display.

Nonvisual NUIs have been developed either to accommodate the abilities of users with visual impairments [23, 12] or to increase available interaction space on mobile devices [28, 8, 17]. Nonvisual NUIs do not rely on the visual sense and are different from recent on-body computing approaches, that use screen-less devices but which project a display on a external surface [22] or the user's skin [11].

Instead, nonvisual NUIs exploit the human ability to interact with their body in an ear and eye free manner using the *proprioceptive* sense [23, 17, 5, 6]. Because nonvisual NUIs keep the user's eyes and ears free, they may be useful in contexts where the use of a screen is considered dangerous, for example, when the user is active, such as walking or driving.

Spatial interaction is an essential component of NUIs, where a touch or a gesture manipulates the state of an on-screen object [8], for example, dragging a file into a folder. Nonvisual NUIs define a virtual display in front of the user, where a user activates a shortcut [17] or drags a virtual object [8, 23] using upper-body gestures. Spatial interaction in nonvisual NUIs is challenging as existing techniques do not provide any spatial feedback [5] and interaction relies entirely on the user's ability to memorize the location of objects. Such interaction may impose a significant cognitive load on the user, especially when multiple objects are present and visuospatial memory tends to fade over time. Though audio coordinates have been explored for spatial feedback [8] this may be difficult to facilitate in mobile contexts due to noise, privacy and social concerns. Studies show that vibrotactile feedback perception does not diminish due to motor activity [26], which makes it an ideal feedback modality for mobile contexts.

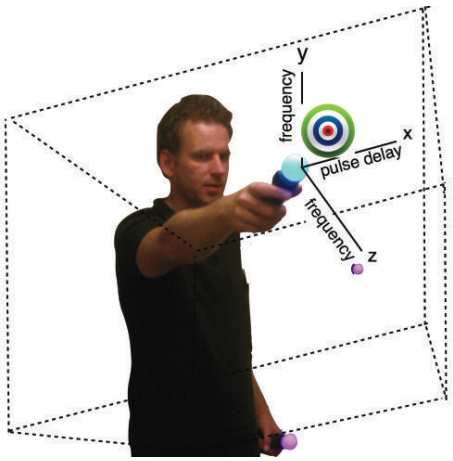

Figure 1: For 3D target selection, vibrotactile feedback guides the user to position their arm such that it touches a nonvisual object defined in a space in front of them, which then allows for manipulating this object using a spatial gesture. The dashed line indicates the frustum of the available search space.

We present a nonvisual display technique that uses haptic feedback to position the user's arm and hand so as to point out the location of a virtual object. The location of a virtual object is then conveyed using the proprioceptive sense, which then enables spatial interaction. Our technique complements existing nonvisual NUIs [8, 23, 17] in that it enables the user to retrieve or to memorize the location of a virtual object without having relying on the user's visual-spatial memory. Our technique is further unique in that target acquisition and spatial interaction are integrated into a single technique that appropriates the user's body for target acquisition as well as target selection, thus creating an *embodied NUI* [4].

Our work significantly extends prior research that has only explored 1D/2D target acquisition using a single arm. In this paper we evaluate the effectiveness of different target scanning techniques for: (1) single arm 3D target acquisition with the goal to facilitate a significantly larger interaction space; and (2) bimanual 2D target scanning, with the goal of improving target selection efficiency. Both studies help further the insight and knowledge of how this non-visual display technique can be used to develop human navigation, rehabilitation and augmented reality/ gaming applications.

*e-mail: fiannaca@cse.unr.edu
†e-mail: morelli@cse.unr.edu
‡e-mail: efolmer@cse.unr.edu

Graphics Interface Conference 2013
29-31 May, Regina, Saskatchewan, Canada
Copyright held by authors. Permission granted to CHCCS/SCDHM to publish in print form, and ACM to publish electronically.

2 Related Work

Research in nonvisual NUIs has initially focused on exploring how touch screen devices can be made accessible to users who are blind, for example, by providing speech feedback when users browse menus [7] or through the definition of custom gestures [12]. Several nonvisual NUIs have been developed for the purpose of increasing available input space of mobile devices without having to compromise their portability. These techniques typically appropriate the device itself into an input device using: (a) its orientation [13], (b) its relative position [10], or (c) gestures made with the device [14]. These techniques only allow for non-spatial interaction, such as scrolling through and activating items from lists.

Virtual shelves [17] is an input technique where users activate shortcuts by positioning a motion sensing controller with an integrated vibrotactor that is tracked using external cameras within a circular hemisphere defined in front of the user. Spatial interaction is limited to activating shortcuts and although users can learn and memorize the location of a particular shortcut using a vibrotactile cue, no spatial feedback is provided. The usefulness of this technique is evaluated with users with visual impairments in a second study [18], which demonstrated that proprioception can be used to create effective nonvisual NUIs.

Gustafson presents an imaginary interface [8] where virtual objects are defined in a plane in front of the user, and whose positions can then be manipulated using spatial gestures. As the name suggests, this interface requires users to memorize the location of virtual objects, which may be challenging to perform, especially when multiple objects are present. An audio-based coordinate system is proposed to retrieve an object's location, but this may be difficult to facilitate in mobile contexts. Familiar interfaces can be used in imaginary interfaces to avoid learning a new interface [9], but spatial interaction is limited to activating shortcuts.

In recent years various on-body computing approaches have been proposed that appropriate arms and hands into input devices [11, 22] but these are all vision based, as they use the user's skin as a display surface by using micro projectors. Recently, several techniques have been explored that appropriate the arm or hand into a non-visual display using a technique called a *tactile-proprioceptive display* [5]. Haptic feedback lends itself well to achieve eye and ear free interaction [27], but haptic feedback on most mobile devices is limited [19] as these typically only feature a single rotary mass motor capable of providing on/off vibrotactile feedback with a fixed frequency and their latency limits the use of sophisticated drive signals [3]. Significantly larger information spaces that are capable of communicating larger amounts and richer types of information to the user can be facilitated through a combination of haptic feedback with proprioceptive information. For example, a navigation tool can be created by having users scan their environment with an orientation aware mobile device where a vibrotactile cue guides the user to point their arm holding the device at a specific point of interest. Target direction is then conveyed to the user using proprioceptive information using their own arm; effectively appropriating the human body into a display. A significant benefit of tactile-proprioceptive displays is that they can be created using hardware that is already present in most mobile devices [5].

Sweep-Shake [28] is a mobile application that points out geolocated information using a tactile-proprioceptive display. The user's location and orientation are determined using a compass and GPS. Vibrotactile feedback that encodes directional information (e.g. directional vibrotactile feedback using pulse delay) renders points of interest. A study with four users found users could locate a 1D target on a 360° horizontal circle in 16.5 seconds. Similarly to Sweep-Shake, PointNav [21] points out geolocated information, but accommodates users with visual impairments.

Ahmaniemi [2] explored target acquisition using a mobile device that consists of a high precision inertial tracker (gyroscope, compass and accelerometer). Directional and nondirectional 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 in 1.8 seconds on average. Target sizes larger than 15° were most effective. Directional feedback was found to be more efficient than nondirectional feedback when target distance is furthest but it negatively affects finding targets that are close.

Previous work: VI Bowling [23] is an exercise game for users who are blind and explores 1D target acquisition and gestured based interaction using a tactile-proprioceptive display. This game was implemented using a motion-sensing controller (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 hit the pins using an upper body gesture that resembles throwing a bowling ball. 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 targets could be found on average in 8.8 seconds and gestures were performed with an aiming error of 9.8°. In subsequent work, we explored 2D target acquisition using one arm [5] . A tactile-proprioceptive display was implemented using a motion-sensing controller, whose position and orientation can be tracked using an external camera and inertial sensing. Its integrated vibrotactor is capable of providing directional vibrotactile feedback using pulse delay and frequency. A tactile-proprioceptive display was implemented whose size was defined by the reach of the user's arm and which defined a rectangular plane in front of the player. Target acquisition was evaluated using an augmented reality space invader game, in which players scan to a random target defined in the display, and shoot it by pulling the controller's trigger. Two different target-scanning strategies were proposed. *Linear* scanning involves finding the target's X-coordinate using an on-target vibrotactile cue, upon which the direction to the target's Y is indicated using frequency modulation. *Multilinear* scanning uses directional vibrotactile feedback that is provided simultaneously on both axes where no pulse delay and maximum frequency indicates the target. A between subjects study with sixteen users found multilinear scanning to be significantly faster than linear scanning. Targets were acquired on average in 7.7 seconds ($SD = 2.8$). Additionally, a second study explored the users' ability to perform spatial gestures by having users touch a target using a thrust gesture. A user study with eight subjects using multilinear scanning found that users could perform a gesture in the direction the controller was pointing with an average aiming error of 20.74°.

3 Study 1: 3D Target Acquisition

Our first study extends prior work on 2D target selection [5] to 3D in order to investigate whether proprioceptive displays can facilitate a significantly larger interaction space. The size of the search space is therefore expanded from a plane to a frustum, whose depth is defined by the length of the user's arm and the location of the camera (see Figure 1) used to track the motion sensing controller that is used to facilitate this display. The back plane has a width that covers the entire horizontal range of the user's arm when it rotates at the shoulder joint and its height is restricted by the camera's resolution. This study is limited to rendering a single target at a time. Based on prior work [5], two different scanning strategies for 3D target acquisition were identified:

- **Multilinear** scanning uses directional vibrotactile feedback on each Cartesian axis of the frustum to indicate the target's location (see Figure 2). In previous work we [5] demonstrated that users were able to scan to a target on two axes simultaneously and we naturally extend this approach to indicate a target's Z-value. Different types of haptic feedback are used on each axis to indicate the direction to the target. The user can find the direction to the target in one gesture by moving

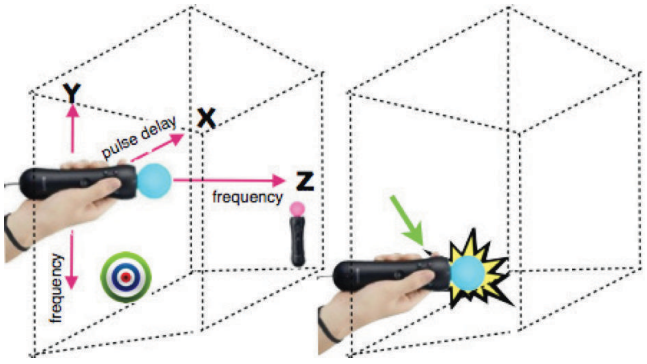

Figure 2: Using multilinear scanning directional haptic feedback provided on the X, Y, and Z-axes guides the user to select the target.

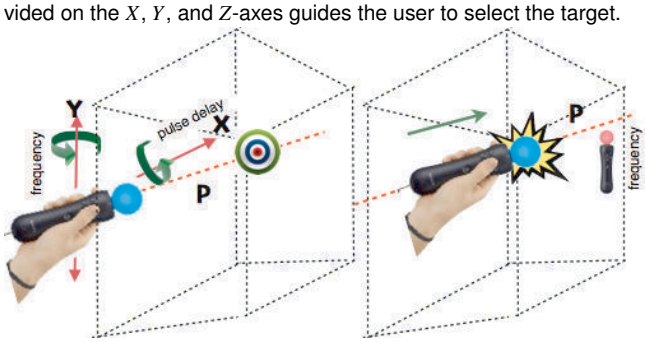

Figure 3: Using projected scanning the user first places the controller outside the frustum and rotates the controller along its X and Y-axes until it points at the target. The user then moves the controller along the projected axis, P, as indicated using directional haptic feedback to select the target.

the controller in any of the 8 directions that lie between the X, Y, and Z-axes. In theory, if the direction to the target on all axes is known, the user can scan directly to the target. This scanning type can be performed regardless of the initial start position of the controller.

- **Projected** scanning is a two-step target acquisition technique. Preliminary experiences with multilinear scanning revealed that scanning along three axes simultaneously was quite challenging to perform and required some amount of practice. To accommodate this limitation, we developed a simpler-to-perform two-step scanning technique. In previous work [5] we also found that subjects were able to perform a directed gesture in the direction their controller was pointing with reasonable accuracy, after finding a target in 2D. Projected scanning is designed on these results and involves performing the following two steps: (1) with the controller initially outside of the frustum users rotate the controller along its own X and Y-axes as indicated using directional vibrotactile feedback until it points at the target; then (2) the user moves the controller along a projected axis (P) that is defined by the controller's elongated shape and its current orientation (see Figure 3). Directional vibrotactile feedback indicates how far to move along the P-axis to select the target. Though projected scanning involves performing two consecutive steps, rotating the controller along its own X and Y-axes may be achieved faster than moving the controller along the coordinate axes of the frustum.

Each of these strategies use one controller for scanning the frustum and one controller for receiving tactile feedback corresponding to the Z or P-position of the controller scanning the frustum. Both

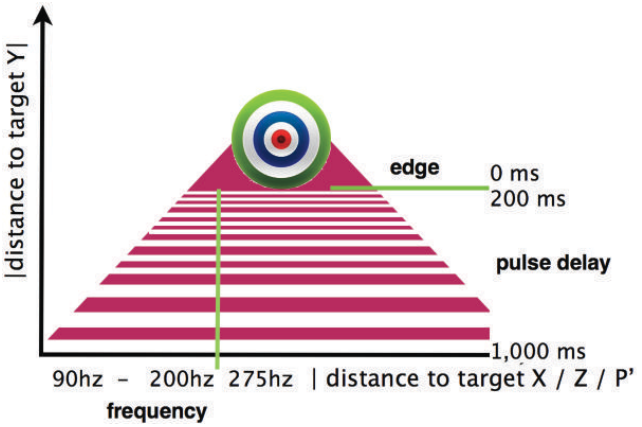

Figure 4: Haptic encoding of directional feedback on each axis, showing how haptic feedback changes in the frustum. When on target pulse delay is zero and frequency is 275Hz. Pulse delay increases linearly from 200ms at the edge of the target to 1.0s (max) at the edge of the frustum. Frequency decreases linearly from 200Hz at the edge of the target to 90Hz at the edge of the frustum.

strategies are therefore equivalent for evaluation. The goal of our user study is to evaluate which of the identified scanning techniques is faster.

3.1 Instrumentation

Our tactile-proprioceptive display is implemented using a commercially available motion sensing controller called the Sony PlayStation Move [1]. The controller's orientation is tracked using inertial sensing. It features an LED that serves as an active marker where the uniform spherical shape and known size of the LED allows for controller's position to be tracked in three dimensions with high precision ($\pm$1.0mm error) using an external camera called the PlayStation Eye, which captures video at 640x480 (60Hz). Directional vibrotactile feedback can be provided using pulse delay or frequency modulation with a range of 91 to 275Hz.

The user scans the frustum with a controller held in their dominant hand, where pulse delay and frequency are used to indicate the direction of the target's X and Y. Use of the dominant hand is motivated by that for motor tasks this yields a better performance than using the non-dominant hand. A Move controller is limited in only being able to provide two types of directional feedback, therefore both scanning techniques use a second controller in the user's non-dominant hand to indicate the target's Z position (multilinear scanning) or its P position (projected scanning) using frequency modulation (see figure 3).

A related study with 1D target selection using a haptic mouse [25] found that targets can be found significantly faster when the difference between the on-the-target cue and close-to-target cue is significantly increased ($\geq$20%) at the border of a target. For target scanning on the Y and Z/P-axes frequency was modulated linearly based on the Y or Z/P distance to target with a maximum value of 200Hz at the edge of the target, which was boosted to 275Hz (maximum) when on target. For the X-axis, the pulse delay was 0ms when on target and 200ms at the edge of the target, which decreased linearly with 3ms/pixel depending on the distance to the target's X-coordinate. The values used in our study were all informed by results from prior studies with tactile-proprioceptive displays [25, 5, 2].

Figure 4 illustrates the haptic encoding scheme for providing directional feedback for the various axes. For multilinear scanning when the user is on the target both controllers provide continuous (pulse delay of 0ms) haptic feedback at 275Hz. For projected scan-

ning when the user points their right controller at the target, this controller provides continuous haptic feedback at 275Hz and when the user selects the target, their other controller will provide haptic feedback at 275Hz.

To compare both scanning types, a simple game was developed that involved destroying targets by selecting them. The faster players destroy a target the more points they score. The use of a game was motived by the fact that games are considered powerful motivators [29], which may allow for measuring optimal performance in a user study. The game runs on a laptop and communicates with a PlayStation 3 to retrieve the current position and orientation of each Move controller and to adjust the vibrotactile feedback. As the controllers are wireless, there is a small latency in our feedback system but we found this lag to be minimal and not to significantly affect our study. To indicate to the player when the controller moves (multilinear) or points (projected) out of the frustum all vibrotactile feedback would be interrupted. Due to the camera's 4:3 aspect ratio, the user is more likely to move or point the controller outside of the Y-range, therefore frequency is used to render the target's Y-coordinate as this provides continuous feedback which makes being or pointing outside of the frustum more noticeable to a user than using pulse delay.

The camera used a 75° field of view with a 640x480 resolution (60fps). The controller's X and Y are reported in pixels and its Z in millimeters. Ahmaniemi [2] found target sizes larger than 15° for a 1D display of 90° to be most effective. For this study a target size of 100 pixels for X, 80 pixels for Y and 100 mm for the P/Z-axis (based on an average arm length of 60cm) was used as to have a similar target size. A single target is defined at a random location within the frustum, excluding a 5% border to avoid scanning too close to the border. The use of random targets as opposed to fixed targets was motivated by the fact that it allows for assessing the user's ability to consecutively scan targets independent of the controller's initial position. Potential applications of our technique, such as an exercise game [24], typically also use random targets. If the controller is within the defined target area for 1s the target is destroyed, a sound effect is played, score is announced, and a new target is generated. Random background music is played to mask the sound of the vibrotactor.

3.2 Participants

We recruited 16 participants (6 female, average age 28.5, $SD = 3.42$). All subjects were right handed and none had any self-reported impairments in tactile perception or motor control. We measured players' height ($M = 170.82$cm, $SD = 7.59$) and arm's length ($M = 59.73$cm, $SD = 2.72$).

3.3 Procedure

Participants were randomly assigned into two eight-person groups (A,B) where group A played the game using multilinear scanning and group B using projected scanning. A between-subjects study design is justified to avoid interference effects, e.g., when users have mastered one scanning technique this may disrupt their ability to learn and use another. User studies took place in a small office room. An observer was present during the study. Participants played the game using their dominant arm while standing. Due to players having different height and arm length, an application that is part of the Move SDK was used to calibrate the position of the player and to define the size of the frustum. Players were placed facing the camera at approximately 8 feet away (recommended optimal distance). Using a visual task that was displayed on the laptop's screen, players would be positioned as such to ensure the full horizontal range of their arm at the shoulder joint would match the horizontal resolution of the camera, e.g., the display ranges 180° by 135° (4:3 aspect ratio). The player would then stretch their arm and press the trigger on the controller to define the frustum's depth.

Table 1: Mean corrected search time (stdev) for each axis (mm/ms)

AXIS	PROJECTED	MULTILINEAR
X	.037 (.022)	.066 (.072)
Y	.044 (.030)	.059 (.058)
P / Z	.034 (.014)	.033 (.026)

Once the position of the player was calibrated a piece of paper was placed under the player and we asked players to keep standing on it while playing the game. The laptop display was turned off to minimize distraction. Players were then instructed what the goal of the game was and how to play the game either using projected or multilinear scanning. Players familiarized themselves with the size of the frustum. For projected scanning, players were instructed to start scanning by placing the controller outside of the frustum, e.g., in front of their body, and rotate the controller to be able to find the direction to the target and then move along the projected axis. For multilinear scanning players were taught how to find the direction to the target on all axes from any starting position. Players played our game briefly until they felt comfortable with scanning targets using their scanning technique. The game would be reset and players would play the game until 20 targets were hit. All targets and all positions and orientations of the controller were recorded in a log file.

3.4 Results

An analysis of collected data reveals significant variance in performance, which reduces after the eight' target. We consider this part of the learning phase and our analysis therefore focuses on the players' performance of acquiring the last twelve targets. The average search time for a target was 12.99s ($SD = 6.90$) for multilinear and 17.46s ($SD = 4.28$) for projected scanning.

Because targets were defined at random, the target distance from the initial start position could vary significantly between trials, though this variation reduces for a larger number of trials. For a fairer comparison, we compare search time corrected for distance. Using the measured user's arm length the target distance on the X and Y-axes can be converted to millimeters, which yields corrected search times of .102mm/ms ($SD = .192$) for multilinear scanning and .075mm/ms ($SD = .095$) for projected scanning. This difference was not statistically significant ($\boldsymbol{t}_{14} = -.769,\ p > .05$). We then analyzed search performance for each axis by calculating the corrected search time based on the last time the target border was crossed. In 14% (projected) and 10% (multilinear) of the targets the player was already within the target range for one specific axis, which resulted in significant outliers in corrected search time for that axis. Table 1 lists the results with the outliers filtered out.

A repeated measures ANOVA found no statistically significant difference between projected and multilinear scanning for corrected search times on all axes ($\boldsymbol{F}_{2,12} = .425$, $p > .05$, Wilk's $\lambda = .904$, partial $\varepsilon^2 = .096$). We then analyzed corrected search times for each axis within each scanning type, but no significant difference between axes for multilinear scanning ($\boldsymbol{F}_{2,21} = .799$, $p > .05$) or projected scanning ($\boldsymbol{F}_{2,21} = .286$, $p > .05$) was found.

For each search we created trace graphs for the controller's position. Figure 5 shows a typical trace for each technique. For multilinear scanning, though users would perform the correct initial motion to find the direction to the target on all three axes (see Figure 5:left), they would typically scan to target's X and Y before scanning to it's Z-coordinate. Only for some of the last targets some users were able to scan to the target on all axes simultaneously. For projected scanning, we found that for larger target distance on the P-axis users would start to deviate on the X and Y-axis as following the projected axis P would become harder (see Figure 5: right).

Logs further show that for multilinear scanning users spent an average of .43s ($SD = .04$) searching for a target outside of the frustum and for projected scanning users pointed their controller out-

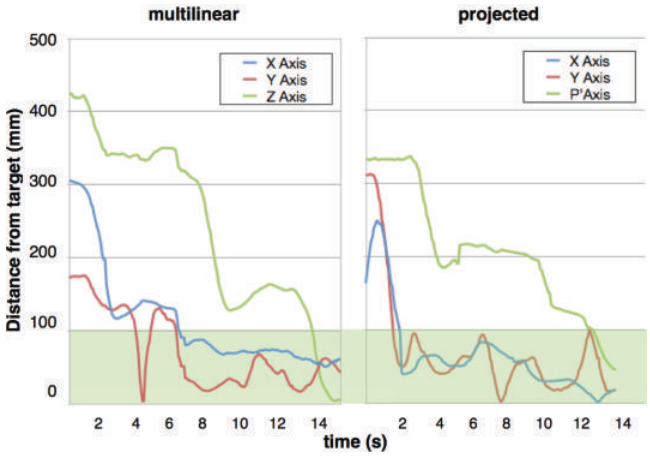

Figure 5: Typical scanning strategies for multilinear scanning (left) and projected scanning (right).

side the frustum on average 2.18s ($SD = 1.48$) per target. This difference was statistically significant ($t_{7.010} = 3.866$, $p < .05$). For projected scanning this data was corrected for when the user starts scanning for the target and the controller is outside the frustum. Closer analysis found that for multilinear scanning this sometimes occurred for targets close to the frustum's edge where users would move the controller through one of frustum's sides when scanning for the target's Z-coordinate. For projected scanning, pointing the controller outside the frustum predominantly occurred on the Y-axis, when users were acquiring the direction to the target.

4 Study 2: Bimanual 2D Target Acquisition

Our second study extends prior work on 2D target selection [5] but extends it such to explore bimanual use. While in the first study, only one of the two controllers was used to scan the search space, in this study we use both controllers for scanning. The goal of this study was to determine whether both arms could be used for target acquisition. Using both arms could possibly allow for faster target acquisition. The size of the search space consists of a vertical plane defined in front of the user, the size of which is determined by the length of the user's arm and the resolution of the camera. Users locate a single target using two controllers. Two scanning strategies for bimanual 2D target acquisition were defined:

- **Split** scanning divides the available search space into two equal sized regions, where each controller implements a display for each region. We use the same haptic encoding scheme as in a previous study [5], i.e., multilinear scanning where different types of haptic feedback modulation are used to indicate the direction to the target on each axis, which allows for the user to search for the target's location on both axes simultaneously.

- **Conjunctional** scanning uses a single display where each controller indicates one of the targets' coordinates using haptic feedback modulation. Users can find the target's X and Y coordinates using one controller for each axis. This is an asymmetric task performed synchronously. Upon finding the coordinates, the target can be selected by moving one controller to the intersection of the found X and Y coordinates.

The choice for these specific scanning strategies was motivated as they each evaluate one potential improvement in performance of bimanual operation. Split scanning may be faster as each controller implements a smaller display that can be scanned through faster. Conjunctional scanning provides insight into whether users can use both controllers at the same time, which may be faster than using a single controller to find both targets' coordinates. Though bimanual operation could allow for multi target scanning, we restrict our study to single targets and the identified strategies are equivalent for evaluation.

Table 2: Average search time (and stdev)

AXIS	SPLIT	CONJUNCTIONAL
Left	6.00 (2.32)	6.72 (2.01)
Right	4.64 (1.92)	7.41 (2.16)
Both	7.07 (1.90)	11.04 (3.15)

4.1 Instrumentation

We used the same setup as for the first user study (see Section 3.1). For split scanning, pulse delay and frequency are used to indicate the direction of the targets X and Y. A short cue indicates in which region the target is rendered. For conjunctional scanning, we use frequency modulation to indicate the target's X-coordinate on the dominant hand controller and the Y-coordinate on the non-dominant hand controller. The same values as for the first study were used for frequency modulation, pulse delay modulation, and the target size. The game for study 1was adapted to facilitate 2D scanning and was used to evaluate both scanning strategies. While targets were defined at random in the first study due to the relative large search space, we defined a grid in which targets appeared. Target locations were counterbalanced using a Latin square where the target order was randomized between trials.

4.2 Participants

We recruited 16 participants (5 female, average age 25.7, $SD = 3.53$). All subjects were right handed and none had any self-reported impairments in tactile perception or motor control. We measured players' height ($M = 1.77.28$cm, $SD = 11.42$) and arm's length ($M = 57.25$cm, $SD = 4.44$).

4.3 Procedure

Participants were randomly assigned into two eight-person groups (A,B) where group A played the game using split scanning and group B using conjunctional scanning. For the same reason as the first study, a between-subjects study design is justified to avoid interference effects, e.g., when users have mastered one scanning technique this may disrupt their ability to learn and use another. We used a similar procedure as in the first study (see Section 3.3). After calibration, both groups received instructions on how to scan for a target using their scanning technique. Players played the game briefly until they felt comfortable performing their scanning technique. The game would then be reset and players would play the game until 20 targets were hit. All targets and all positions and orientations of the controller were recorded in a log file.

4.4 Results

The average target search time was 7.07s ($SD = 1.90$) for split scanning (A) and 11.04s ($SD = 3.15$) (B) for conjunctional scanning. This difference was statistically significant ($t_{14} = -2.854$, $p < .05$). Unlike the first study, we did not analyze the search time corrected for distance, since for split scanning subjects would often lower the hand holding the controller that was not active. Upon become active this would lead to very large distances causing an unfair comparison between scanning techniques. Table 2 lists the target search time per hand for each technique. For conjunctional scanning, only a few users were able to scan with both controllers along the axes at the same time, where the rest would scan for the coordinates sequentially. For conjunctional scanning, we therefore calculate search time from the moment the user begins scanning with that controller. No significant difference in search time was found between hands for split scanning ($t_{14} = -0.657$, $p > .05$) or

conjunctional scanning ($t_{14} = 1.276$, $p > .05$) demonstrating that users were just as proficient with either hand. Logs further show that for split scanning users spent an average of .45s ($SD = .32$) searching for a target outside of the display and for projected scanning this was 2.59s ($SD = 1.65$) per target. This difference was statistically significant ($t_{7.51} = -3.363$, $p < .05$).

5 Discussion and Future Work

Both studies show that tactile-proprioceptive displays are not superfast, but they allow for communicating a type of information, i.e., a 2D/3D point in a space in front of a user, with a significantly large spatial resolution that would otherwise be difficult to communicate using conventional types of haptic feedback, such as tactons.

Our first user study revealed no significant difference in performance between multilinear and projected scanning, which contradicts our previous study with 2D scanning [5]. Our study identified the advantages and disadvantages of using each scanning technique.

Projected scanning allows for more quickly finding the direction to the target (as rotating the controller is faster than moving the controller in the frustum), but users spend significantly more time searching for the target outside of the frustum as with multilinear scanning moving the controller within the frustum is physically constrained by the user's arm and therefore less likely to happen. Multilinear scanning allows a user to scan to the target directly, but in our study we rarely observed users being able to do this and instead they followed a two-step process similar to projected scanning where users first acquired X and Y simultaneously and then proceeded scanning the target's Z.

Similar to preliminary experiences, for some users, scanning along three axes simultaneously turned out to be too difficult to perform, which could indicate that we have run into a human limitation, as this was easier to perform for 2D scanning [5]. On the other hand a few users were observed to be able to do this for the last targets, which could indicate that it could also be a matter of practice. Due to a limitation of the Sony Move controller to only be capable of providing two types of vibrotactile feedback (pulse delay and frequency), we were required to use a second controller in the other hand to indicate the target's Z-coordinate. Users may have found it difficult to combine and interpret stimuli from both hands into a single sensation. If a third type of directional vibrotactile feedback is used, i.e., amplitude, simultaneous provision of three types of haptic feedback using a single device will introduce the effect of frequency being perceived as more dominant and this typically drowns out amplitude perception [20]. For 3D target selection using two controllers may actually be the most optimal, as this interference problem will not occur.

For targets defined close to the frustum's back plane, projected scanning seemed more difficult to perform as users would more easily deviate from the projected axis, which often requires the user to move the controller outside the frustum as to reacquire the target. The length of the user's arm and the resolution of the camera define the size of the frustum. As a result the search space on the X and Y-axes is almost twice the size of the search space on the Z-axis, which does not really allow for a fair comparison of search performance between axes. For such a comparison, a uniform search space would be more suitable, but then users are more likely to move or point outside of the frustum. Though the arm length between users did not vary significantly, the volume of the frustum defined by the user's arm may vary significantly.

Our second study showed that the average search time for split scanning was 7.07s. In prior research [5] we found a scanning time of 7.7s, which demonstrates that using both hands is slightly faster. These findings are consistent with previous work on bimanual use of pointing devices [16], though the performance gain we found is much smaller. To an extent this is explained by the fact that users would typically lower the arm that was not actively scanning for a target, so it took longer to find a target, as it took some time for the user to raise their arm again. For conjunctional scanning, only a few users were able to scan with both controllers along the axes at the same time. This could have been a matter of convenience, though users were demonstrated how to do this, or it could indicate that this was very challenging to perform, and that users would require more practice to master this.

Reflecting on all studies with target acquisition using tactile-proprioceptive displays in 1D, 2D and 3D space, we observe the following. Ahmaniemi [2] found an average target search time of 1.8 seconds for a 90° 1D display. In previous work we [24] found a scanning time of 7.7 seconds for a 180° by 135° display. In our 3D scanning study we found a search time of 12.99 seconds (multilinear) for a 180° by 135° by arm length display. Extrapolating these results to match the size of each display, we can observe that search time nearly doubles each time an axis is added to the search space, e.g., $3.6s(x) \rightarrow 7.7s(x,y) \rightarrow 12.99s(x,y,z)$. However, because axes were not exactly of equal sizes between studies this finding should be further substantiated in subsequent research.

Our target selection studies were constrained to conveying a single target at a time, though for some applications, such as exergames, we may require to render multiple targets at the same time, as to stimulate larger physical activity. For 3D target selection, rendering multiple targets is limited by technical constraints of the controller used. For 2D target selection users should be able to use both controllers to select two targets using split scanning. For an exergame, using this technique you could simulate punching targets in 2.5D where targets are defined on the surface of a sphere whose size is defined by the length of the user's arm.

Our tactile-proprioceptive display relies on an external vision system to determine the 3D position of the user's controllers. For mobile contexts, where ear and eye free interaction is most useful we can imagine that our display can be implemented using a wearable camera. Recent advances in 3D cameras may allow for the user to wear a small camera on their chest, which allows for accurate arm tracking where directional haptic feedback can be provided using a miniature haptic device [15]. This approach is different from how we evaluated our display as targets were defined in the frame of a fixed camera. Using a wearable camera, targets will be relative to a user and may be subject to interference from walking and moving.

Potential Applications: In addition to complementing existing nonvisual mobile spatial interfaces [8, 17], useful applications of our technique could include developing whole-body exercise games for individuals who are blind [23], as this typically involves punching and kicking virtual targets that are defined in a 3D space around the player. Though scanning for a 3D target with the arm stretched is some form of physical activity, it is unlikely to engage a player into levels of physical activity that are high enough to be considered healthy. Targets could be defined in 2D and the size of the display could be reduced to allow for rapid gestures. Alternatively, a rehabilitation or yoga like game could be facilitated using our technique where finding 3D targets using both arms would force the user into a particular position, e.g., both arms extended to the user's sides.

Another application area of our technique could be human navigation systems. Several tactile-proprioceptive techniques have already been developed that use the users arm to point out the direction towards an object of interest [28, 21] but they do not tell the user how far away the object of interest is. Our technique could enhance these existing techniques by using the Z-coordinate of a target's location to convey the relative distance to the point of interest. For example, if the user has to stretch their arm completely to touch the target's Z-coordinate this could indicate that the object of interest is 10 km away and if it is close to the user's body 1 km. This allows for intuitively finding objects without requiring to look at a display or using audio, which could be useful, for example, to develop a search and rescue application.

6 Conclusion

This paper presents a nonvisual display technique that addresses a limitation of existing nonvisual NUIs; as for spatial interaction these techniques rely on the user's ability to memorize the location of virtual objects, which may be challengingf. Our technique uses haptic feedback to position the user's arm and hand as such to point out the location of a virtual object. The location of a virtual object is then conveyed using the proprioceptive sense, which then enables spatial interaction. A user study with 16 participants evaluates two different single-arm target- scanning strategies for selecting an object in 3D. A second study with 16 participants evaluates two bimanual target-scanning strategies for 2D object selection. Both studies provide useful insights in the effectiveness of tactile-proprioceptive displays, which could help develop new applications based on this technique.

Acknowledgements

This work is supported by NSF Grant IIS-1118074

References

[1] Sony Playstation Move controller, `http://us.playstation.com/ps3/playstation-move/`, access date: 10-10-2010.

[2] T. T. Ahmaniemi and V. T. Lantz. Augmented reality target finding based on tactile cues. In *Proceedings of the 2009 international conference on Multimodal interfaces*, ICMI-MLMI '09, pages 335–342, New York, NY, USA, 2009. ACM.

[3] L. M. Brown and T. Kaaresoja. Feel who's talking: using tactons for mobile phone alerts. In *CHI '06 extended abstracts on Human factors in computing systems*, CHI EA '06, pages 604–609, New York, NY, USA, 2006. ACM.

[4] P. Dourish. *Where the Action Is: The Foundations of Embodied Interaction (Bradford Books)*. The MIT Press, 2004.

[5] E. Folmer and T. Morelli. Spatial gestures using a tactile-proprioceptive display. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 139–142, New York, NY, USA, 2012. ACM.

[6] E. Gardner and J. Martin. *Principles of Neural Science, fourth edition.*, chapter Coding of Sensory Information, pages 411–429. Number 21. McGraw-Hill, 2000.

[7] T. Guerreiro, P. Lagoa, H. Nicolau, D. Gonalves, and J. Jorge. From tapping to touching: Making touch screens accessible to blind users. *Multimedia, IEEE*, 15(4):48 –50, oct.-dec. 2008.

[8] S. Gustafson, D. Bierwirth, and P. Baudisch. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 3–12, New York, NY, USA, 2010. ACM.

[9] S. Gustafson, C. Holz, and P. Baudisch. Imaginary phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 283–292, New York, NY, USA, 2011. ACM.

[10] C. Harrison and S. E. Hudson. Minput: enabling interaction on small mobile devices with high-precision, low-cost, multipoint optical tracking. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1661–1664, New York, NY, USA, 2010. ACM.

[11] C. Harrison, D. Tan, and D. Morris. Skinput: appropriating the body as an input surface. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 453–462, New York, NY, USA, 2010. ACM.

[12] 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 of the 10th international ACM SIGACCESS conference on Computers and accessibility*, Assets '08, pages 73–80, New York, NY, USA, 2008. ACM.

[13] J. Keränen, J. Bergman, and J. Kauko. Gravity sphere: gestural audio-tactile interface for mobile music exploration. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1531–1534, New York, NY, USA, 2009. ACM.

[14] J. Kim, J. He, K. Lyons, and T. Starner. The gesture watch: A wireless contact-free gesture based wrist interface. In *ISWC '07: Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, pages 1–8, Washington, DC, USA, 2007. IEEE Computer Society.

[15] K. Kim, J. E. Colgate, M. A. Peshkin, J. J. Santos-Munné, and A. Makhlin. A miniature tactor design for upper extremity prosthesis. In *Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies*, FBIT '07, pages 537–542, Washington, DC, USA, 2007. IEEE Computer Society.

[16] C. Latulipe, C. S. Kaplan, and C. L. A. Clarke. Bimanual and unimanual image alignment: an evaluation of mouse-based techniques. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, UIST '05, pages 123–131, New York, NY, USA, 2005. ACM.

[17] F. C. Y. Li, D. Dearman, and K. N. Truong. Virtual shelves: interactions with orientation aware devices. In *UIST '09: Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 125–128, New York, NY, USA, 2009. ACM.

[18] F. C. Y. Li, D. Dearman, and K. N. Truong. Leveraging proprioception to make mobile phones more accessible to users with visual impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '10, pages 187–194, New York, NY, USA, 2010. ACM.

[19] J. Luk, J. Pasquero, S. Little, K. MacLean, V. Levesque, and V. Hayward. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 171–180, New York, NY, USA, 2006. ACM.

[20] K. Maclean and M. Enriquez. Perceptual design of haptic icons. In *In Proceedings of Eurohaptics*, pages 351–363, 2003.

[21] C. Magnusson, M. Molina, K. Rassmus-Gröhn, and D. Szymczak. Pointing for non-visual orientation and navigation. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 735–738, New York, NY, USA, 2010. ACM.

[22] P. Mistry and P. Maes. Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, SIGGRAPH ASIA '09, pages 11:1–11:1, New York, NY, USA, 2009. ACM.

[23] T. Morelli, J. Foley, and E. Folmer. VI-Bowling: a tactile spatial exergame for individuals with visual impairments. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS'10)*, pages 179–186, Orlando, Florida, USA, 2010.

[24] T. Morelli and E. Folmer. Twuist: A discrete tactile-proprioceptive display for eye and ear free output on mobile devices. In *Proceedings of Haptics Symposium 2012 (HAPTICS'12)*, pages 443–450, Vancouver, Canada, 2012.

[25] 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):993–1004, 2007.

[26] L. J. Post, I. C. Zompa, and C. E. Chapman. Perception of vibrotactile stimuli during motor activity in human subjects. *Exp Brain Res*, 100(1):107–20, 1994.

[27] 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.

[28] S. Robinson, P. Eslambolchilar, and M. Jones. Sweep-shake: finding digital resources in physical environments. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 12:1–12:10, New York, NY, USA, 2009. ACM.

[29] K. Squire. Changing the game: What happens when video games enter the classroom? *Innovate: Journal of Online Education*, 1(6), 2005.

[30] D. Wigdor and D. Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2011.