Speech-Filtered Bubble Ray: Improving Target Acquisition on Display Walls

Edward Tse, Mark Hancock, Saul Greenberg
Interactions Laboratory, University of Calgary
2500 University Dr. N.W., Calgary, Alberta, Canada T2N 1N4
[tsee, msh, saul]@cpsc.ucalgary.ca

ABSTRACT

The rapid development of large interactive wall displays has been accompanied by research on methods that allow people to interact with the display at a distance. The basic method for target acquisition is by *ray casting* a cursor from one's pointing finger or hand position; the problem is that selection is slow and error-prone with small targets. A better method is the *bubble cursor* that resizes the cursor's activation area to effectively enlarge the target size. The catch is that this technique's effectiveness depends on the proximity of surrounding targets: while beneficial in sparse spaces, it is less so when targets are densely packed together. Our method is the *speech-filtered bubble ray* that uses speech to transform a dense target space into a sparse one. Our strategy builds on what people already do: people pointing to distant objects in a physical workspace typically disambiguate their choice through speech. For example, a person could point to a stack of books and say "the green one". Gesture indicates the approximate location for the search, and speech 'filters' unrelated books from the search. Our technique works the same way; a person specifies a property of the desired object, and only the location of objects matching that property trigger the bubble size. In a controlled evaluation, people were faster and preferred using the speech-filtered bubble ray over the standard bubble ray and ray casting approach.

Categories and Subject Descriptors

H5.2 [**Information interfaces and presentation**]: User Interfaces. – Interaction Styles.

General Terms

Design, Human Factors

Keywords

Large display walls, speech, gestures, speech filtering, multimodal, freehand interaction, pointing

1. INTRODUCTION

The recent development of large high-resolution wall display technology has been accompanied by parallel developments of suitable interaction techniques. Most systems still use input controls that are separate (vs. direct touch) from the wall, e.g., multiple mice [17], game consoles, or remote controls that require a person to navigate through items via buttons. Touch-sensitive surfaces work well when people can reach the wall, e.g., Smart Technologies Interactive Whiteboards (smarttech.com) and the MERL DiamondTouch digital table [5].

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICMI'07, November 12–15, 2007, Nagoya, Aichi, Japan.
Copyright 2007 ACM 978-1-59593-817-6/07/0011...$5.00.

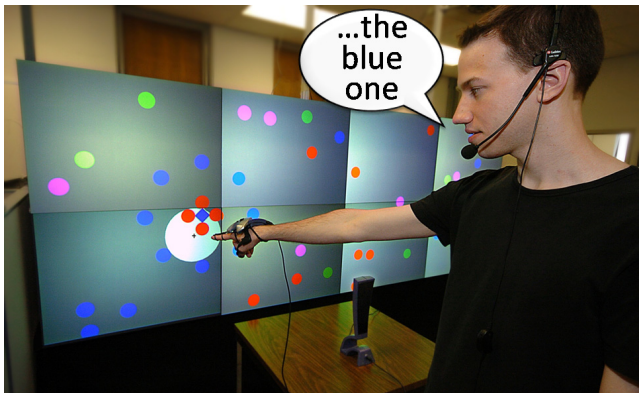


Figure 1. The wall study hardware configuration

Yet in many large display settings, some tasks are best performed from a distance. For example, in everyday conversations people may use nearby display walls to view and interact with content relevant to their discussion. They may find it inconvenient or interruptive to approach the display or acquire specialized input devices. If the display wall is large, some regions of the screen may not be easy to reach, e.g., the upper region may be out of reach for some people. The display wall's position within a furnished environment also impacts a person's proximity to it, e.g., when mounted out of direct reach in a public place like an airport or restaurant, or as a common information wall situated in a control room where operators are seated at workstations.

The general problem is: how can people effectively select items with large displays from a distance? As we will see in §2, a variety of strategies have been developed by others. Some move distant items closer to where the person is actually working on the display screen [1,2]. Most others use variants of *ray casting*, where a person's pointing action is interpreted as a 'ray' hitting the screen. Ray casting is of particular interest to us. It is the most natural for people to do, and it also serves as a gesture that is easily understood by others involved in the activity.

While ray casting is reasonable for large targets, it is slow and error prone when people try to select small targets on the display. Indeed, we believe target acquisition will become a serious issue both as distances between people and the display increase, and as improved screen and input resolutions create more available pixels per inch. Fitts Law partially predicts this problem (see §2), but the situation is exacerbated by the natural shake in people's hands when pointing [13], and by inaccuracies in ray casting input technologies. Even a very small shake when pointing translates to large jitters over a distance of several feet.

Tse, E., Hancock, M. and Greenberg, S. (2007) Speech-Filtered Bubble Ray: Improving Target Acquisition on Display Walls. Proc 9th Int'l Conf. Multimodal Interfaces (ACM ICMI'07), (Nov. 12-15, Nagoya, Japan). ACM Press.

One way to get around this is by increasing the apparent size of small targets. In particular, Grossman et al. [7] developed the *bubble cursor* to simplify pointing in a sparse environment; the cursor is surrounded by a 'bubble' resized to envelope the closest target. This technique works well if the surrounding space is fairly sparse since the bubble can grow reasonably large, but is ineffective in dense spaces as the bubble has little room to grow.

Thus the specific problem addressed by this paper is: can we improve target acquisition via ray casting, even when targets are densely packed together?

We have been inspired by observations of everyday communication: people often roughly point to an area containing several objects, and then use speech to discriminate the particular object of interest within that zone. For example, one might point to a coat rack containing several coats and ask "please, pass the red one". In this case, speech helps the listener filter out coats that are not red from the range of possible targets being pointing to by the speaker. We claim that an analogous multimodal speech and pointing system, which we call *speech-filtered bubble ray* (or speech bubble for short), can help people select targets on a large interactive digital wall from several feet away, as illustrated in Figure 1. A person specifies a property of the desired object, and only the location of objects matching that property triggers the bubble ray size.

2. RELATED WORK

There are two areas of relevant related work: input techniques that improve target acquisition by 'optimizing' Fitts Law parameters, and input techniques for distant freehand pointing.

2.1 Optimizing Fitts Law Parameters

Fitts Law is commonly used to model target acquisition [10]. The Shannon formulation of Fitts Law [11] states that the movement time (*MT*) that it takes to acquire a target of width *W* and distance (or amplitude) *D* is predicted by:

$$MT = a + b\log_2\left(\frac{D}{W}+1\right)$$

a and *b* are empirically determined constants, and the logarithmic term is called the *index of difficulty* (*ID*). The equation predicts that smaller target widths and larger distances (from the current location) will increase selection time. Thus target selection can be improved by decreasing *D*, by increasing *W,* or both.

Decreasing Target Distance (*D*): Baudisch et al. [1] reduce target distance by bringing distant targets closer to the user. Their *Drag-and-Pop* method analyzes the directional movements of the cursor, and then brings virtual proxies of the potential targets towards the cursor (e.g., a folder or application). Studies of drag-and-pop showed selection to be faster for large target distances. However, their method cannot determine when the user intended to select distant targets versus those nearby. Thus the presence of distant objects can make selection difficult for nearby targets.

Bezerianos et al.'s *Vacuum* method [2] is somewhat similar, but also allows the user to control the angle of distant targets that they were interested in and supports multiple object selection. Selection time was found to be similar for single targets but significantly faster for multiple target selection.

Increasing Target Width (*W*): Kabbash and Buxton [8] increased the target width by increasing the cursor size. Instead of a single pixel hotspot as seen in standard cursors, area cursors have a larger active region for selection. By setting *W* to be the width of the area cursor, they showed that selection of a single pixel target could be accurately modeled using Fitts Law. Thus, very small targets are easier to acquire. However, area cursors are problematic in dense target spaces where multiple targets could be contained in an area cursor.

McGuffin and Balakrishnan [12] increased the target size dynamically as the cursor approached. They found that users were able to benefit from the larger target width even when expansion occurred after 90% of the distance to the target was traveled. They also showed that overall performance could be measured with Fitts Law by setting the width to the size of the expanding target.

Increasing *W* and Decreasing *D*: A different approach dynamically adjusts the control-display gain (*C:D*). By increasing the gain (cursor speed) when approaching a target and decreasing it while inside a target the motor space distance and width are decreased and increased, respectively. Blanch et al. [3] showed that performance could be modeled using Fitts Law, based on the resulting larger *W* and smaller *D* in motor space. However, problems arise when there are multiple targets, as each slows down the cursor as one travels towards it.

As mentioned, Grossman et al. developed the Bubble Cursor to ease target acquisition in a sparse display [7]. The cursor is surrounded by a dynamically resizing bubble so that only the closest target is enveloped by the bubble. An example is shown in Figure 2 (left): the bubble around the cross hair cursor expands until it just touches the nearest target. This effectively increases target width (since the bubble gets bigger), and decreases target distance (because less distance needs to be traveled to reach the target). The problem is that other nearby targets, called *distracters*, limit the size of the bubble. For example, if the four objects surrounding the cursor in Figure 2 (left) were closer together, the bubble would be much smaller. In other words, the width of the target is dependent on the distance of the closest distracters adjacent to it, as it expands so that only the closest target is selected at any time. This new target size is called the *Effective Width* (*EW*). Their study shows that Bubble Cursor's performance can be modeled using Fitts Law by setting $W = EW$.

2.2 Freehand Pointing at Large Displays

Ray Casting is a commonly used technique for pointing to distant objects on a large display (e.g., [4, 12, 14, 18]), where the cursor is drawn as the intersection of the ray from the hand/pointer and the screen. Laser pointers are an obvious candidate for implementing ray casting, and many people have explored how they can be implemented and used. For example, Myers et al. [13] considered different laser pointer form factors (a pen, a laser pointer mounted on a glove, a scanner, and a toy gun) to see how they minimized hand jitter and affected aiming. Parker et al.'s *TractorBeam* [14] affords selection on a tabletop display by having people point the tip of the six degree-of-freedom pen at distant targets. Other ray casting devices include data gloves, wands tracked by motion capture systems, and so on.

Improving Ray Casting. The basic selection methods described in §2.1 can be applied to ray casting. For example, the *TractorBeam* [15] includes: Expand Cursor (cursor expands when close to the target), Expand Target (target expands when cursor approaches) and Snap-to-Target (cursor jumps to the closest target). The Snap-to-Target proved quickest but had a high error

rate. Expand Cursor was slowest and proved problematic when multiple targets were nearby, but had the lowest error rate.

Selection. While target acquisition is all about pointing, a complementary act is to indicate the actual act of selection. In a conventional computer, the mouse may move the cursor, but it is the button-down operation that 'selects' the object under the cursor. In distant pointing, a common approach is to use a selection button or similar control onto the pointing device, e.g., [13]. Vogel and Balakrishnan [19] present two gesture-based selection methods applied to distant freehand pointing for large high-resolution displays. With the *thumb trigger* the user presses their thumb against their hand to perform a selection. With the *air tap*, the user moves their index finger to indicate selection.

Multimodal input. Early attempts at distant freehand pointing include Bolt's [4] Put-That-There multimodal system where individuals could interact with a large display via speech commands qualified by deictic reference, e.g., "Put that…" (points to item) "there…" (points to location). These systems, however, still use basic ray casting for target acquisition. Kaiser *et al.* [2003] fused the n-best recognition results of speech, gesture, and gaze input to disambiguate the object that a person was pointing to. Studies revealed that this approach was effective for handling recognition errors and uncertainty in the recognition of any one modality.

3. Speech-Filtered Bubble Ray

Many of the techniques described in §2, whether for direct touch, separate input controls, or ray casting variants, attempt to 'optimize' Fitts Law performance by decreasing target distance or increasing target width. However, in a dense space, nearby distracters (i.e., other potential targets) limit the effectiveness of these approaches; in the worst case, they degrade to simple ray casting. Yet in practice, dense information-rich spaces are more typical than sparse ones, e.g., an icon-filled desktop, a web page or document filled with text, a control panel comprising many widgets (buttons, sliders, etc.), or a highly populated node and link graph. The difficulty of target acquisition via free-hand pointing is onerous at a distance due to pointing inaccuracies of the input device and hand shake. The question is: how can we improve target acquisition in such a densely packed space?

As introduced earlier, people already use speech to qualify pointing gestures made with one's hands. Mary might point to a region on a book shelf with her finger and say "the green one", or Fred might point to a file browser on a large display and say "the Latex file" (Figure 1). Viewers interpret the gesture as indicating the rough regions they should be attending, and then use the speech act to decide upon the object of interest. Both speech and gesture work in concert; neither provides enough information by itself to discriminate the object of interest.

Our new interaction method works on the same principle. First, we adapted the bubble cursor to work with freehand ray-casting (*vs.* a mouse) for distant selection – we call this the *bubble ray*, but it is otherwise identical to the bubble cursor. Second, we added speech-filtering capabilities to it to create the *speech-filtered bubble ray*. As a person moves the bubble ray towards an object, he or she simultaneously uses speech to inform the system of a particular property of that object. At that point, objects that don't have that property are filtered from consideration by the bubble cursor algorithm. Unless objects with the same property are very close to one another, the effect is that speech filtering makes densely packed target spaces sparser. Sense Shapes by Kaiser *et al.* [9] projected a fixed 3D cone from a person's hand while this cone dynamically expands and contracts in the Speech Bubble Ray.

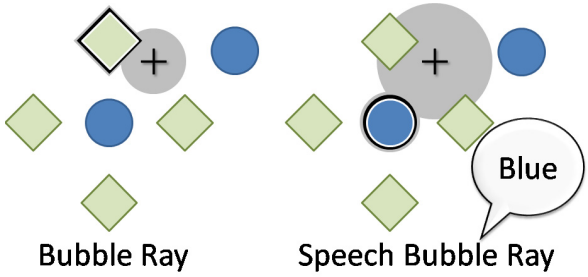


Figure 2. Bubble Cursor and Speech Bubble Ray methods

To illustrate this by example, consider Figure 2, which shows a standard bubble ray implemented as a bubble ray on the left and our speech-filtered bubble ray on the right. Say the person wants to select the blue circular object in the middle, and she is moving the cursor towards it from the top. If the person does not say anything, the system behaves as a normal bubble ray, where the bubble expands to touch and select the closest target. As Figure 2 (left) shows, the bubble is somewhat small because other distracter targets are nearby. However if the person says "the blue one", the bubble ray selection space is filtered to include only blue objects and to ignore the green diamonds, and the bubble ray expands accordingly (Figure 2 right, also Figure 1). Here, the bubble's size is constrained by the next closest blue object. Comparing the two figures illustrates that even though the blue objects are not that far apart, the bubble is considerably larger and thus target acquisition is much easier for the speech bubble. Selection can then be performed with techniques such as a button press on a control held in the non-dominant hand, or through a gesture [19]. By simply saying "none" or "cancel" the user can remove the speech filter.

We stress that the speech qualifier does not perform any selection. We chose this design as there may be many objects near the target that share a similar property, and an incorrect selection could be made if speech filtering also performed a selection. Thus, it may be more beneficial to speak the speech qualifier early in the target acquisition, to maximize the effect of the filter, e.g., before or during the ballistic move just as one visually tracks the object and recognizes one of its properties.

By filtering according to speech commands, our technique increases the effective width of targets. Fitts Law [4] predicts that this increase will reduce movement times. However, the technique has an added cognitive overhead of deciding on what to filter and an added time to actually speak a command. Consequently, we performed an empirical study that compares the performance cost of our speech-filtered bubble ray to two other distant freehand pointing techniques. These include ray casting (commonly used as a control in such studies) and the bubble ray.

4. User Study

Our study goal was to compare selection times and error rates between three pointing techniques: ray casting, bubble ray and speech-filtered bubble ray.

4.1 Method

4.1.1 Participants

Thirty students (17 male, 13 female) from a local university participated in the study. Ages ranged from 18-34; all participants

were right-handed and daily computer users with normal or corrected to normal vision. Half of all participants (15) had experience with large display pointing devices (e.g. Smart Board, Nintendo Wii). When asked to rate their English fluency on a scale of 1 to 5 (5 being completely fluent), 27 participants rated themselves as 5, one rated herself as 4, one rated himself as 3, and one rated himself as 2. When asked about what language they primarily spoke at home, 21 reported English with the remaining 9 reporting other languages – Asian, Middle Eastern and French.

4.1.2 Freehand Pointing and Selection Techniques

We adapted three selection techniques for use in our environment.

Ray casting is implemented as a crosshair cursor that represents the intersection of a ray starting from a person's hand and intersecting with the display wall. As the person moves their hand, the corresponding crosshair also moves.

Bubble ray adapts Grossman's bubble cursor technique [7] for distant freehand pointing on a wall display. Our version differs only in that the person uses ray casting instead of a mouse cursor. The bubble around the crosshair dynamically expands so that only the nearest target is enveloped in the bubble (Figure 2, left). We use the formula described in [7] to control the bubble's size. We note that bubble cursor as applied to ray casting has not been evaluated elsewhere, but we expect that the performance can be modeled (as it is with the mouse in [7]) using Fitts Law by setting the target width (W) to the effective width.

Speech-filtered bubble ray (or **speech bubble** for short) is visually identical to bubble ray for the purposes of the experiment, except that the bubble size is adjusted to the filtered objects. This is actually a 'worst case' visualization, as in practice filtered targets could be faded to emphasize the sparseness of the new selection space. Using a microphone headset, people spoke a single property of the target (its color) into a speech recognizer to activate the filter.

For all three techniques, if the crosshair or bubble is within a target's active region, the target is highlighted with a white and black border (see Figure 2). This emphasis is visually similar to the underlining of links on a web page or the blue highlight seen with the single-click icon selection mode in Microsoft Windows XP. This is especially important for both bubble ray approaches, as it emphasizes that a target has been acquired and that there is no need to further move the cursor closer to the target.

4.1.3 Apparatus

As seen in Figure 1, we used a 2.94 m x 1.10 m display surface composed of eight modular ambient display (MAD) boxes [16] each containing a 1024 x 768 LCD projector for a total resolution of 4096 x 1536. All projectors were connected to a single workstation with two Matrox QID Pro display adapters that each support four displays. Our system is designed in C++ using a large OpenGL window spanning across eight displays.

For input we used a six degree-of-freedom Essential Reality P5 Data Glove, a low cost input device intended for computer gaming. We used only the x and y values for our experiment, thus the position of the cursor was only affected by the position of the glove relative to the sensor. Tilting the hand would not change the position of the cursor.

The data glove sensor was placed at the bottom-centre and 0.83 m in front of the wall. Participants were asked to stand in a square marked by masking tape 1.80 m in front of the wall, shifted 0.83 m to the left of centre, so that the right arm of the participant was aligned at the centre of the screen. Freehand pointing was performed with the right arm.

Participants used a Labtec LVA 7330 noise-cancelling microphone for the speech bubble technique. Because we did not want speech recognition errors to influence our results, we used a Wizard of Oz speech recognition technique: the target colour was activated when any speech was recognized. If the participant said the wrong target colour, the experimenter would mark the trial as having a speech error.

We gave participants a wireless slide remote to perform selections in the non-dominant hand. We preferred this to a selection technique in the dominant hand to minimize any drift from the intended selection location.

As a side note, we expect that future vision and audio processing systems can easily detect user actions without the need for specialized glove tracking devices and headsets.

4.1.4 Task

For each trial, participants were presented with a screen full of distracters coloured red, green, blue, and pink, visible in Figure 1. One target was presented in a different shape than the rest (either a diamond or a circle). Participants were asked to select the differently-shaped target as quickly and as accurately as possible. Once selected, the screen would refresh with a new set of distracters and a new target. Thus, the *colour* variable was used for speech filtering and the *shape* variable for target identification.

4.1.5 Design and Procedure

We used a repeated measures within-participant factorial design. Our independent variables were:

- *technique* (ray casting, bubble ray, speech filtering)
- 6 *distracter layouts* as configured in different inner and outer ring widths, and which affect how large the bubble ray and speech bubble ray can grow (Figures 3 and 4).

The distracter layouts need explanation. The six distracter layouts are a combination of two factors: inner width and outer width (Figure 3). These two factors could not be considered separately, because the outer width was constrained to be no smaller than the inner width (and thus, they are not independent). *Inner ring* consists of distracters coloured differently than the target: thus its distance from target restricts bubble size only in the bubble ray condition. *Outer ring* consists of distracters of the same colour as the target, thus its distance restricts the bubble size in the speech bubble condition, once the speech command has been spoken. For example, in the small inner / large outer ring width condition (Figure 4, top-right), the bubble ray is constrained by the small inner ring of distracters (left bubble) and the speech bubble is constrained by the large outer ring of distracters (right bubble).

Both the inner and outer widths vary from small (6.5 cm from target centre), to medium (14.1 cm), and large (24.7 cm). The small size is typical of targets stacked side-by-side (e.g., lines in a text document), the middle is similar to the separation of file icons in a folder, and the large size represents a sparse space on a desktop.

Figure 4 shows the six combinations of inner and outer widths. For each condition, the minimum size of the bubble ray is shown on the left, and the minimum size of the speech bubble (once a colour has been spoken) is shown on the right. For the conditions where the inner and outer widths are the same (top-left, bottom-left, and bottom-right), only one ring of distracters of the same

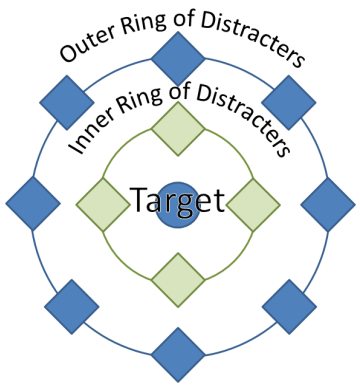


Figure 3. Distracter layout

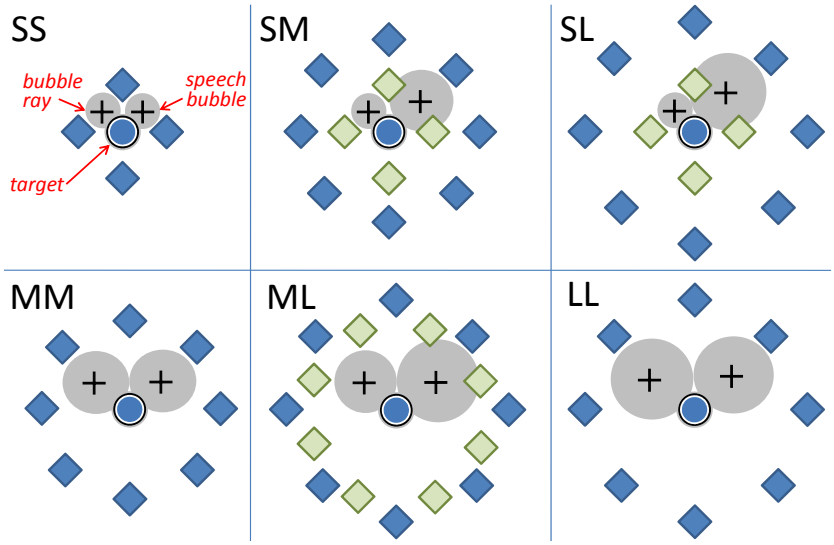


Figure 4. The six distracter layout conditions

colour as the target need be shown (as this ring is sufficient to limit the size of the bubble in both the bubble ray and speech bubble conditions). We will refer to these six conditions as: Small inner ring / Small outer ring (SS), SM, SL, MM, ML, and LL.

We kept constant the distance to the next target (87.5 cm), the diameter of the targets (6.4 cm), the number of non-overlapping distracter targets placed randomly around the screen (74), and the number of possible target colours (4: red, green, blue, pink). All targets had a circular activation area regardless of the shape shown on the screen (diamond or circle).

Participants completed each technique and the distracter layout combinations six times, for a minimum of 108 trials per participant. If a participant made an error during a trial, either by selecting the wrong target or saying the wrong speech command, the trial was repeated. A brief sound cue would indicate if the correct or incorrect selection was made.

Presentation of the three techniques was counter-balanced using a Latin Square. The experiment consisted of 3 blocks (one per technique), with each block following the procedure of:

- 36 practice trials
- 36 trials
- Incorrect trials repeated
- Questionnaire (what did you like/dislike about the technique?)

The practice trials repeated exactly the same conditions seen in the experiment. Each block of 36 trials was randomized.

Participants were asked to complete a post-test questionnaire asking them to compare each of the three techniques after the experiment.

4.1.6 Hypotheses

We had the following hypotheses for this experiment:

H1: The speed of selection will not vary for ray casting.

H2: Bubble ray will be faster in proportion to the inner ring width.

H3: Speech bubble will be faster in proportion to the outer ring width.

H4: Bubble ray will be faster and result in fewer errors than ray casting when the inner ring width is either medium or large.

H5: Speech bubble will be faster and result in fewer errors than ray casting when the outer ring width is either medium or large.

H6: Speech bubble will be faster and result in fewer errors than bubble ray when the outer width is larger than the inner width.

H7: Bubble ray will be faster than speech bubble when the inner ring width and outer ring width are the same.

We hypothesize H7 because of the added overhead in speech bubble of both determining and speaking the command.

4.1.7 Data Collection

During the experiment we logged the position of the cursor, the time, speech volume, and the closest target every 10 milliseconds. When a selection was made we recorded the total trial time, any selection or speech errors (marked by the experimenter) and recorded the positions and colours of every target on the screen.

5. Results & Discussion

To analyse our data, we performed a 6 (distracter layout) × 3 (technique) within-participants ANOVA. We used the average over the six repetitions of both target selection time and number of errors (either incorrect spoken command or missed targets) as dependent measures. We performed the same two analyses with the additional between-participants factor of gender and found no additional main effects or interactions. We present the two-way ANOVA for simplicity.

5.1 Speed

There was a main effect of technique ($F(2,58) = 29.5$, $p < .001$). Post-hoc comparisons showed that all pairwise differences were significant ($p < .01$) and that participants selected targets the fastest with the speech bubble ($M = 2.62$ s, $SD = 0.09$ s) followed by the bubble ray ($M = 2.97$ s, $SD = 0.09$ s), and the ray casting technique was slowest ($M = 3.42$ s, $SD = 0.12$ s).

There was a main effect of distracter layout ($F(5,145) = 47.0$, $p < .001$). There was also significant interaction between distracter layout and technique ($F(10,290) = 10.1$, $p < .001$). While we expected the former main effect (changing target size should affect speed of target acquisition), we are most interested in how these changes affect each of the techniques differently. Thus, we will only discuss this latter interaction. We present pairwise differences broken down both by technique and by distracter layout, as they are both illustrative.

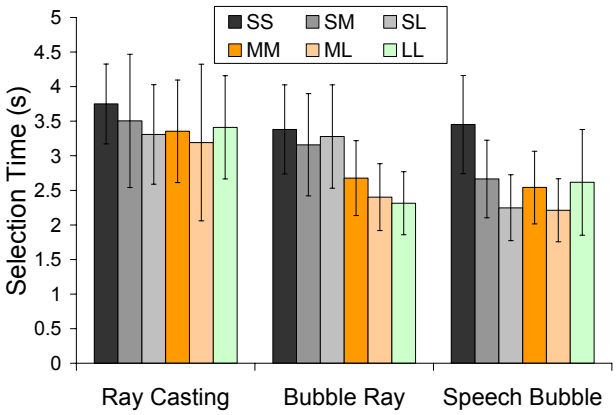


Figure 5. Target selection times for distracter layout, separated by technique.

	Ray Casting	Bubble Ray	Speech Bubble
SS vs. SM	$p = 0.11$	$p = 0.1$	**$p < .001$**
SS vs. SL	**$p < .001$**	$p = 0.37$	**$p < .001$**
SS vs. MM	**$p < .01$**	**$p < .001$**	**$p < .001$**
SS vs. ML	**$p < .01$**	**$p < .001$**	**$p < .001$**
SS vs. LL	**$p < .01$**	**$p < .001$**	**$p < .001$**
SM vs. SL	$p = 0.20$	$p = 0.36$	**$p < .001$**
SM vs. MM	$p = 0.35$	**$p < .001$**	$p = 0.14$
SM vs. ML	$p = 0.06$	**$p < .001$**	**$p < .001$**
SM vs. LL	$p = 0.5$	**$p < .001$**	$p = 0.65$
SL vs. MM	$p = 0.76$	**$p < .001$**	**$p < .001$**
SL vs. ML	$p = 0.50$	**$p < .001$**	$p = 0.55$
SL vs. LL	$p = 0.35$	**$p < .001$**	**$p < .01$**
MM vs. ML	$p = 0.39$	**$p < .001$**	**$p < .001$**
MM vs. LL	$p = 0.65$	**$p < .001$**	$p = 0.41$
ML vs. LL	$p = 0.18$	$p = 0.21$	**$p < .001$**

Table 1. Distracter layout time differences separated by technique. Pairwise significance values are in bold.

Figure 5 shows the target selection times for each distracter layout separated by technique. Table 1 shows significant pairwise differences for distracter layout pair. These pairwise differences partially confirm hypotheses H1, H2, and H3. We found no significant differences in selection times for the ray casting technique (H1) with the exception of the SS condition being slower than the rest. This exception is likely due to the fact that, in the SS condition, the pattern of the single ring of targets is smaller as a whole than in any other condition, making it more difficult to recognize the target before acquiring it and thus increasing cognitive load. In the bubble ray condition, the smallest inner width was slower to select than larger inner widths, confirming H2. The MM condition was also slower than the LL condition, as H2 predicts, however, the ML condition was unexpectedly faster than the MM condition. We suspect this exception is again due to the fact that, in the ML condition, the distracter layout of surrounding targets improved the participants' ability to recognize the location of the center target, artificially improving selection time for this condition. In the speech bubble condition, the targets with a small outer width were slowest to select (H3), the targets with a medium outer width were also slower to select than those with a large outer width (H3) with the exception of the LL condition. Again, the visual cues provided as a side effect of our setup may have been the cause of this exception. In the LL condition, there is only one ring of distracter targets, distant from the actual target, making the pattern of targets more difficult to recognize.

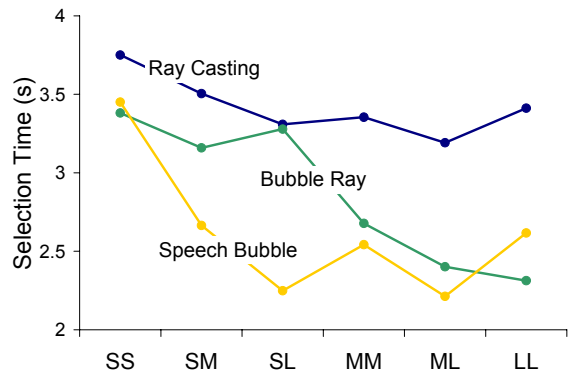


Figure 6. Target selection times for each technique, separated by distracter layout.

	Ray Casting vs. Bubble Ray	Ray Casting vs. Speech Bubble	Bubble Ray vs. Speech Bubble
SS	**$p = .02$**	**$p = .04$**	$p = .60$
SM	$p = .08$	**$p < .001$**	**$p < .001$**
SL	$p = .83$	**$p < .001$**	**$p < .001$**
MM	**$p < .001$**	**$p < .001$**	$p = .26$
ML	**$p < .001$**	**$p < .001$**	**$p = .02$**
LL	**$p < .001$**	**$p < .001$**	**$p = .04$**

Table 2. Time Differences between techniques separated by distracter layout. Pairwise significance values are in bold.

Figure 5 shows the target selection times for each technique separated by distracter layout. Table 2 shows the significant pairwise differences for each. These pairwise differences confirm hypotheses H4, H5, and H6. As H4 predicts, the bubble ray technique was significantly faster than ray casting whenever the inner ring width was medium or large (MM, ML, LL). As H5 predicts, the speech bubble technique was significantly faster than ray casting whenever the outer ring width was medium or large (SM, SL, MM, ML, LL). As H6 predicts, the speech bubble technique was faster than the bubble ray technique whenever the outer ring width was larger than the inner ring width (SM, SL, ML). As H7 predicts, speech bubble was significantly slower than bubble ray in the LL condition, likely due to the overhead required in speaking the command. In addition to our predicted results, we found that ray casting was significantly slower than both bubble ray and speech bubble in the SS condition.

5.2 Error

The average number of errors for any trial was 0.7 (SD = 1.0). Due to the small number of errors, not much can be read from these differences. However, some of these differences were statistically significant. There was a main effect of technique ($F(2,58) = 5.7$, $p < .01$). Post-hoc comparison revealed that participants performed significantly more errors with ray casting than with speech bubble ($p < .01$). There was no significant difference between bubble ray and either ray casting ($p = .07$) or speech bubble ($p = .19$). There was a main effect of distracter

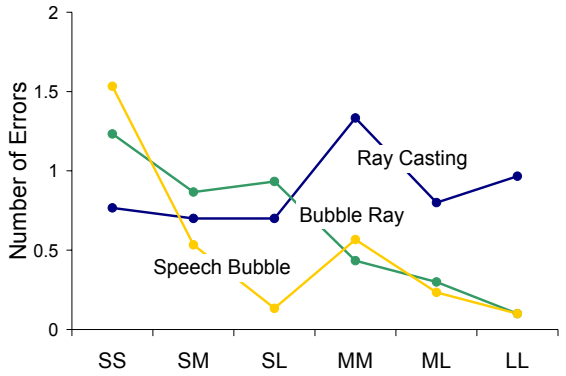


Figure 7. Number of errors for each technique, separated by distracter layout.

	Ray Casting vs. Bubble Ray	Ray Casting vs. Speech Bubble	Bubble Ray vs. Speech Bubble
SS	$p = .13$	$p = .03$	$p = .44$
SM	$p = .55$	$p = .49$	$p = .13$
SL	$p = .26$	$\mathbf{p < .01}$	$\mathbf{p < .001}$
MM	$\mathbf{p < .01}$	$\mathbf{p < .01}$	$p = .40$
ML	$\mathbf{p = .03}$	$\mathbf{p < .01}$	$p = .54$
LL	$\mathbf{p < .001}$	$\mathbf{p < .001}$	$p = 1.00$

Table 3. Error differences between techniques, separated by distracter layout, pairwise significance values in bold.

layout ($F(5,145) = 10.6$, $p < .001$) and an interaction between distracter layout and technique ($F(10,290) = 5.6$, $p < .001$). We will again only discuss the interaction.

Figure 7 shows the number of errors for each technique separated by distracter layout. Table 3 shows the significant pairwise differences for each. These pairwise differences further support our results for speed and confirm hypotheses H4, H5, and H6. For H4, the bubble ray resulted in significantly fewer errors than ray casting when the inner width was medium or large (MM, ML, LL). For H5, speech bubble resulted in significantly fewer errors than ray casting when the outer width was medium or large (SL, MM, ML, LL). This trend also existed for the SM condition, but was not significant. For H6, speech bubble resulted in significantly fewer errors than bubble ray when the difference between inner and outer widths was the largest (SL), but this difference was not significant for the SM or ML conditions.

5.3 Questionnaires

Participant's post-test questionnaire responses revealed a preference for speech bubble as the most liked and easiest technique to use.

Figure 8 shows participant responses to the most liked and most disliked method: 18 liked speech bubble the most, 9 chose bubble ray, and 3 chose ray casting. When asked about which technique they most disliked the opposite effect was observed: ray casting was chosen by 20 participants, 7 disliked bubble ray and 3 disliked speech bubble. Participants' comments reflected their selections as one participant wrote that speech bubble "...makes it easier to select the different shapes by filtering color" while ray casting "was the least forgiving". A few disliked the bubble ray technique saying "I didn't like how the bubble changes in size" and "the jittering of the size of the bubble became a distraction". This problem is caused by the natural shake in people's hands and is further exacerbated by the noise present in the input device.

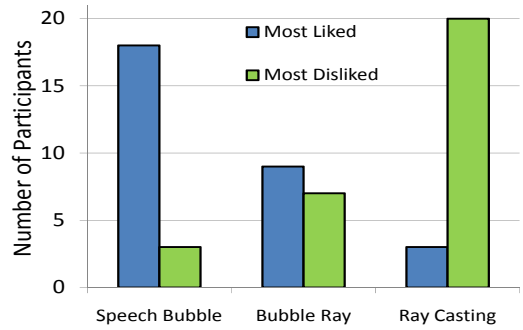


Figure 8. Participant responses of most liked and most disliked technique

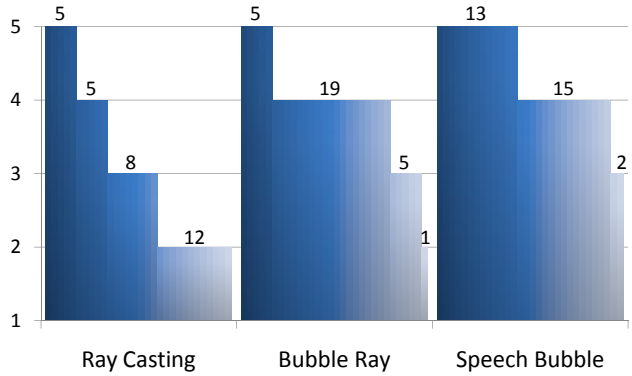


Figure 9. Individual participant breakdown to "I found technique easy to use", 5-strongly agree, 1-strongly disagree

Participants were also asked to say if they agreed that each technique was easy to use on a 5-point Likert scale (1 being strongly disagree, 5 being strongly agree) As seen by the area covered by each technique in Figure 9, speech bubble was ranked higher ($M = 4.4$, $SD = 0.6$) than bubble ray ($M = 3.9$, $SD = 0.7$) and ray casting ($M = 3.1$, $SD = 1.1$). For the speech bubble condition, all participants ranked its ease of use as either strongly agree, agree or neutral. Ray casting had the highest percentage of neutral and disagree responses (20 of 30 participants).

For the people who did not favour the speech bubble, some stated that ray casting was easy to use because it was "like a mouse pointer" and thus most closely matched their everyday use of a computer mouse. Others preferred ray casting and bubble ray over the speech bubble because they "didn't have to speak". Some participants also found wearing a headset microphone uncomfortable.

Language and gender deserve mention. We examined the most liked preferences broken down by gender and language spoken at home but did not notice any notable effects. Some non-native English speakers commented that "coordination between speech and pointing of objects was a bit confusing / delaying".

5.4 Overall Discussion

Our results show that the speech bubble technique provides the performance gain that we had expected and that speech bubble is preferred by most people. Specifically, all our hypotheses were confirmed, suggesting that speech filtering can benefit target selection by effectively increasing target width, even for densely-packed targets. In particular, the speech bubble technique performed as well or better than ray casting and bubble ray in most cases. The only exception was the large inner / large outer

width condition, suggesting that the added overhead of speaking the command becomes slightly detrimental when the surrounding targets are very sparse and speech filtering provides no expected benefit. However, this degradation is only to a level slightly less than bubble ray, but is still faster and less error prone than ray casting. In practice, a person might simply choose not to speak the property, and performance would resort to using bubble ray (the default behaviour when no filter is spoken). In all other conditions, the overhead of speaking was negligible, and was far outweighed by the benefit of filtering.

We also mentioned several other techniques in §2 whose performance is compromised by nearby distracters [1,2,3,8,11]. We believe that speech filtering could be applied to these techniques as well, with performance gains similar to our speech bubble.

6. DESIGN CHALLENGES

A challenge in designing interactions that leverage speech-filtered bubble ray is that the system must reveal properties of the targets that need to be selected. Of course, for bubble ray, the position of every target must be revealed. For speech bubble, the properties of the target used for filtering must also be revealed. For example, in a large desktop environment, the system would need to reveal/detect the position, colour and names of targets if those properties were to be used for speech filtering.

As with bubble cursor, the visual distraction of the bubble can hinder performance when there are very few targets and the bubble grows to become larger than the size of the screen. This problem is exacerbated when speech filtering is used to filter dense target spaces. To correct for this problem, a size limit can be placed on the bubble to limit the visual distraction. Alternatively, a gradient could be used to fade the bubble to transparent past a fixed size, so that the bubble can be much larger with less visual distraction.

7. CONCLUSION

We introduced the *speech-filtered bubble ray*, a technique for improving target acquisition using distant freehand pointing on large display walls by using properties of the target to filter the selection space. While Fitts Law suggests that performance of this method is better than a standard bubble ray or ray casting, speech filtering does incur some cost as people need to determine the target property to filter and say it out loud. Our empirical results provide evidence that the benefits of speech filtering (even when additional visual effects are omitted) significantly outweigh these costs, and effectively make dense target spaces sparser. That is, this multimodal interaction improves selection performance from a distance over large digital displays.

Our future work includes several threads: First, we will investigate the use of speech-filtered bubble ray for interacting with existing applications designed for a keyboard and a mouse. For example, how effective is speech bubble when used for selecting links in a web browser when targets are rectangular in shape? Second, we will see how speech filtering can be combined with whole-hand selection techniques (e.g., thumb trigger [19]) or speech selection (e.g., "here" [4]). Third, we will investigate speech filtering in a multiple-user setting over different display orientations such as a large digital table or a combination of wall and table displays. Finally, we will apply speech filtering to see how it can improve other selection methods compromised by nearby distracters [1,2,3,8,11].

8. REFERENCES

[1] Baudisch, P., Cutrell, E., Robbins, D., Czerwinski, M., Tandler, P., Bederson, B., and Zierlinger, A. (2003). Drag-and-Pop and drag-and-pick: Techniques for accessing remote screen control on touch and pen operated systems, Proc. Interact, 57-64.

[2] Bezerianos, A., Balakrishnan, R. (2005) The Vacuum: Facilitating the manipulation of distant objects. Proc. CHI 2005, ACM Press, 361-370.

[3] Blanch, R., Guiard, Y., and Beaudoin-Lafon, M. (2004) Semantic pointing: improving target acquisition with control-display ratio adaptation. Proc. ACM CHI '04, 519-525.

[4] Bolt, R.A., Put-that-there: Voice and gesture at the graphics interface. *Proc ACM Conf. Computer Graphics and Interactive Techniques Seattle*, 1980, 262-270.

[5] Dietz, P., Leigh, D., (2001) DiamondTouch: A Multi-User Touch Technology, Proc. UIST '01, ACM Press, 219-226.

[6] Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 47, 181-196.

[7] Grossman, T., Balakrishnan, R. (2005) The Bubble Cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. Proc. CHI '05, 281-290.

[8] Kabbash, P. and Buxton, W. (1995) The "Prince" technique: Fitts' law and selection using area cursors. Proc. ACM CHI '95, 273-279.

[9] Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. (2003) Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. Proc. ICMI '03, ACM Press, 12-19.

[10] MacKenzie, I.S. (1989). A note on the information theoretic basis for Fitts' Law. *Journal of Motor Behavior*, 21:323–330.

[11] Mackenzie, I. S. (1995) Movement time prediction in human-computer interfaces. In R. M. Baecker, W. A. S. Buxton, J. Grudin, and S. Greenberg, editors, *Readings in Human-Computer Interaction*. Kaufmann, second edition.

[12] McGuffin, M., Balakrishnan, R. (2005) Fitts' law and expanding targets: Experimental studies and designs for user interfaces. ACM TOCHI, 12(4), ACM Press, 388-422.

[13] Myers, B., Bhatnagar, R., Nichols, J., Peck, C., Kong, D., Miller, R., and Long, C. (2002) Interacting At a Distance: Measuring the Performance of Laser Pointers and Other Devices. Proc CHI'02, 33-40.

[14] Oviatt, S. (1997) Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction* 12.

[15] Parker, K., Mandryk, R., Nunes, M., Inkpen, K. (2005) TractorBeam Selection Aids: Improving Target Acquisition for Pointing Input on Tabletop Displays. Proc. Interact '05,

[16] Schmidt, R., Penner, E., Carpendale, M. S. T. (2004) Reconfigurable Displays. Workshop on Ubiquitous Display Environments; at UBICOMP 2004. ACM Press,

[17] Tse, E. and Greenberg, S. (2004) Rapidly Prototyping Single Display Groupware through the SDG Toolkit, Proc. Australasian User Interface Conference, Australian Computer Society Inc., p101-110.

[18] Tse, E., Shen, C., Greenberg, S., Forlines, C. (2007) How Pairs Interact Over a Multimodal Digital Table, Proc. ACM CHI '07.

[19] Vogel, D., Balakrishnan, R. (2005). Distant freehand pointing and clicking on very large high resolution displays. Proc. ACM UIST 2005, 33-42.

Addendum: poster presented at confernece

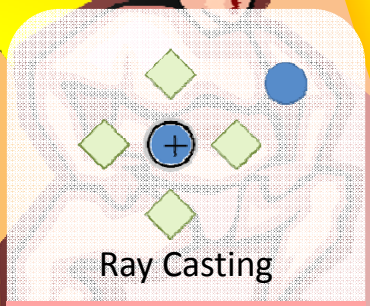
UNIVERSITY OF CALGARY

Speech Filtered Bubble Ray

Improving Target Acquisition on Display Walls

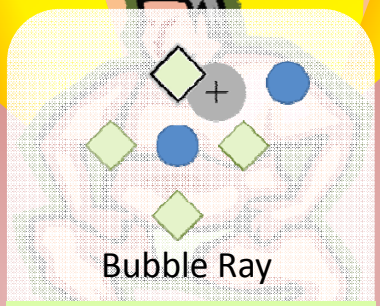
Edward Tse, Mark Hancock, Saul Greenberg

3 CONTENDERS



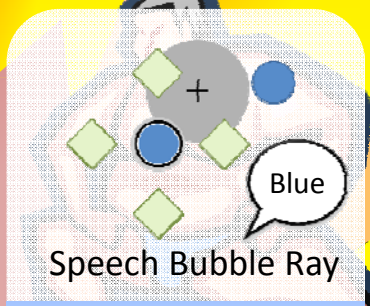
Ray Casting

Must point within the target to select



Bubble Ray

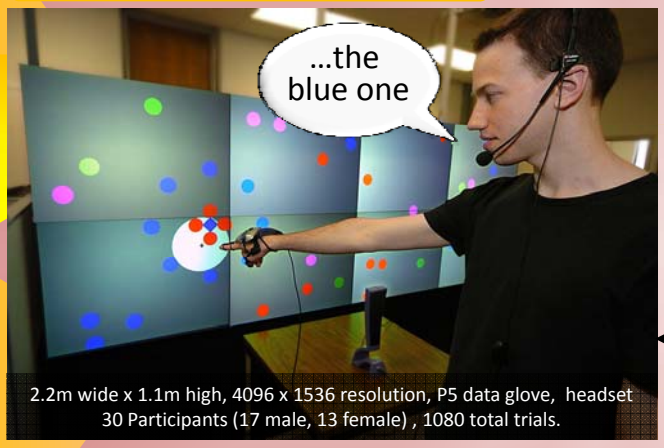
Nearest target is always selected



Speech Bubble Ray

Speech filters the nearest target selection

THE CHALLENGE

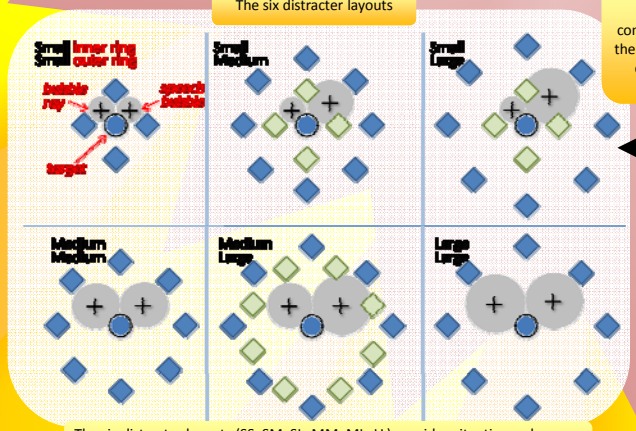


2.2m wide x 1.1m high, 4096 x 1536 resolution, P5 data glove, headset
30 Participants (17 male, 13 female) , 1080 total trials.

Participants were asked to select targets on a large wall display as quickly and as accurately as possible. They performed selections over six different distracter configurations (described below).

DISTRACTER LAYOUT

The six distracter layouts



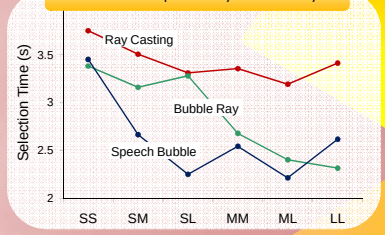
The six distracter layouts (SS, SM, SL, MM, ML, LL) provides situations where ray casting is as effective as bubble ray and speech bubble (e.g. SS).

In our experiment the inner ring (different colour) constrains bubble ray while the outer ring (same colour) constrains the speech bubble size.

THE RESULTS

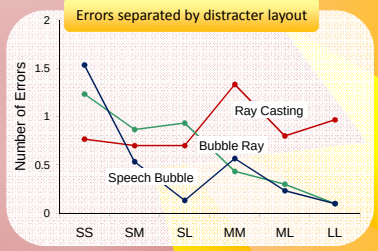
ROUND 1 SPEED

Selection time separated by distracter layout



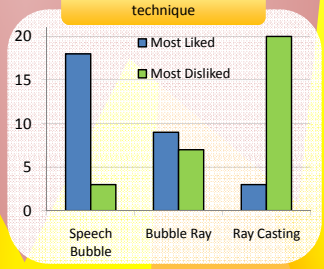
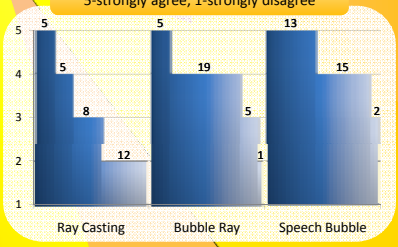
ROUND 2 ACCURACY

Errors separated by distracter layout



ROUND 3 PREFERENCE

Most liked and disliked technique



I found this technique easy to use.
5-strongly agree, 1-strongly disagree