SUDOA Report

Jonas Nockert, nockert@kth.se
(supervised by Christopher Peters)

May 25, 2018

Summary

A non-technical overview of Orbits[1] and Pursuits[2], both based on the same technique which enables calibration-free interaction with graphical devices using only gaze.

Orbits has been developed with small devices, like smart watches, in mind and Pursuits more oriented towards larger displays and public interaction. A number of potential applications have been described in previous papers[3] so instead of reiterating these, I will discuss my idea of using this technique as a possible means to provide interfaces for young children before they can use hands and fingers to interact with electronic devices.

[1] See Esteves et al. (2015c), presented at the 28th Annual ACM Symposium on User Interface Software & Technology 2015 where it won a best paper award, Esteves et al. (2015a), presented at the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, and also the related talk by Esteves et al. (2015d) (video).

[2] See Vidal et al. (2015) and the related video by Vidal (2013).

[3] See Vidal et al. (2013), Vidal et al. (2015), and Pfeuffer et al. (2013).

Background

Gaze tracking

Gaze tracking tries to establish where one is looking[4], e.g. fixated on a point in space or tracking a moving object, by taking into account relative pupil position, head position, etc. Usually this is done with specialized head- or screen-mounted hardware but if gaze-tracking is to become more common, it is likely that sensors will have to be integrated in the devices users would like to use hands-free. Lacking special eye-tracking hardware, many devices today include video calling capapabilites where the video camera can be used as a sensor (albeit usually not with the same precision).

[4] Here differentiated from eye tracking where we might be interested in many other aspects of the movement of the eye such as acceleration.

As the eyes are a relatively small feature of the face, device-mounted eye-trackers need very good accuracy if they should be consistently capable to register minute changes in gaze. At a distance of half a meter from a display, shifting our eyes as little as one degree represents about one centimeter of gaze shift on the display. The Tobii Pro X3–120, a very capable screen-mounted eye tracker, can generally pick up eye movements as small as half a degree under good conditions[5].

[5] See *Tobii Pro X3-120 Accuracy and precision Test report* 2015.

However, in order to reach this kind of precision, eye tracking sensors need to be calibrated to each user's eyes. Recalibration is often also needed when lighting conditions change, especially if moving from indoors to outdoors. This step could make it cumbersome to just pick up a device and use it. For new users, it could even be difficult to know when and how to calibrate.

Gaze interaction

With regular graphical user interfaces, we often use *affordances*[6]

[6] Physical or visual properties of an interactive element guiding the user to its use, e.g. a flip button has two states and only one of the states could be active at a time.

from the physical world to guide their use. Gaze interaction has no counterpart in this sense. How do we even know that an interface supports gaze interaction in the first place? What does a well-designed, intuitive gaze-operated interface element resemble? Do we "click" these elements by using our eyelids somehow or just stare at them long enough?[7]

[7] Picture a dialog box with the question "Really delete file?" and two gaze elements representing yes and cancel in order to get a sense of the potential pitfalls if the interaction model is not thought through.

Now, many of the mobile devices we use today are relatively small and this means that their graphical interface elements are even smaller. Which, in turn, means gaze tracking has to be of very high accuracy if we want to be able to discern exactly which element is being interacted with. The physical size of the devices, as well as their often limited battery and cpu capacity, makes integrating good gaze-tracking difficult. At the same time, it is exactly the small size of devices that make them so suitable for gaze interaction.

On the other side of the spectrum, larger screens and public use makes it difficult to provide good interaction as screens are often too big or too far away for touch, and mice and other pointing devices are seldom a good fit. Here we can use good eye tracking hardware and controlled lighting but calibration poses a problem.

Pursuits

Pursuits introduces a new type of graphical user interface element that is based on movement (see figure 1). A user can select an element by following its specific movements.

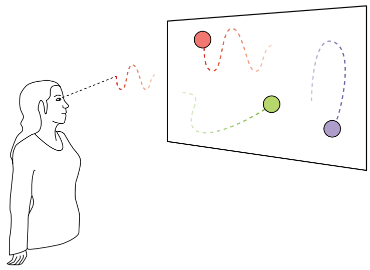

Figure 1: Some examples of Pursuits gaze patterns (Vidal et al., 2015, p. 8).

What makes Pursuits especially interesting is that is based on the smooth pursuits movements of the eye, which is a type of movement that only happens when we are following something with our eyes. In fact, most people can't reproduce this movement on their own, which means that triggering false positives while "just looking" can largely be avoided.

As this technique does not depend on having to identify *where* on the screen a user is gazing, only that the gaze is moving in a specific pattern, it seems to be less dependent on exact readings and, better yet, calibration is not necessary as only relative eye movements are relied upon.

Smooth pursuits is a precise eye movement with little lag, so it enables the use of multiple controls which can be separated by pattern, movement offset, and/or velocity. Even if we would decide to use only horizontal, sinusoidal controls, at least eight simultaneous controls can still be supported[8]. With multiple patterns, we are not limited in terms of controls but rather by the usability aspects of having that many moving controls visible on the screen at one and the same time.

[8] Vidal et al., 2015, p. 10.

Despite initial concerns with moving interactive elements, especially in terms of affordance, Pursuits uses movement to its advantage: "While HCI research has attended to target movement as a problem [...], Pursuits embraces it and allows for dynamic interfaces which naturally attract the eyes."[9] That is, given that people will generally follow a moving object without thinking about it, we could have the system visually respond to this interaction in order for users to notice and be able to use it without instructions[10] (*Hello! Looking at me? Follow my movements to...*)

[9] Vidal et al., 2013, p. 440.

[10] Ibid., p. 440.

Orbits

The wrist watch makes it possible to determine the time with just a flick of the wrist and a short glance. It is this interaction that Orbits is modelled after.

Most interactions with (digital) watches have so far been limited to buttons and the crown. Modern smart watches have extended this to wrist movements and touch displays. Since the watch face is tiny, even with perfect motor skills in our arms and fingers, we can't just bring interactions from larger mobile devices since the interactive elements are so small. If a user has less than perfect mobility, it is even more difficult.

Orbits uses the same basic technique as Pursuits but present specific controls based on circular gaze patterns:

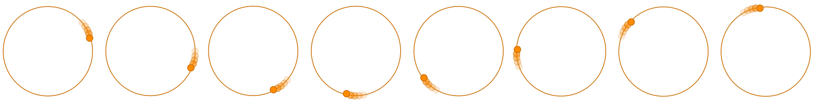

Figure 2: This figure shows some frames of a clockwise orbit (with a speed of 120°/*s*, it will complete a revolution every three seconds). See Esteves et al., 2015b for a video that gives a much better sense of the idea.

Orbits allow for both discrete and continuous controls. One can select an orbit by following its trajectory for a short while (less than a second is usually enough) and an action that is tied to its selection can be triggered, e.g. answering a call or start playing a song. One could also let the following gradually change something, e.g. the longer you follow the trajectory, the louder the volume.

The larger the orbit control is, the easier it is to pick up that someone is following it, which is expected. When it comes to rotation speed though, the paper describes testing four alternatives; 60, 120, 180, and 240 degrees per second, and finding that 120°/*s* gave the best results, noting that this matches the range of speed smooth pursuits operate in. One can imagine that the optimal rotational speed can be further fine-tuned.

Discussion

By actually implementing different applications, both for small and large displays, we know Pursuits and Orbits work not only in theory but in practice. However, one important caveat is that Orbits used a head-mounted eye tracker so some kind of technological leap needs to happen until we can have Orbits on our smart watches and mobiles.

That is, unless the technique works using built-in cameras. I would imagine the moving controls would have to be rather large but it would be very interesting to see if it could work using today's technology. It would then gives us access to cheap yet precise enough, calibration-free, gaze interactions for practically all mobile devices, including laptops, until proper gaze tracking becomes ubiquitious.

Having moving graphical user interface elements is definitely not optimal for most interactions but despite the limitations, I think it provides a lot of room for innovation precisely because it is so simple and we humans interact so naturally with motion.

Applications

It is not difficult to come up with a number of use cases[11] where a user can not or does not want to use their hands and fingers to interact with a device.[12] However, I imagine that once we get gaze interaction in the hands of users, all kinds of use cases we did not at all anticipate for gaze interaction will surface.

[11] For some examples of potential applications, see Vidal et al., 2013; Vidal et al., 2015; Pfeuffer et al., 2013.

[12] Likewise, it is not difficult to identify that gaze interaction can exclude visually impaired users if we do not provide alternatives. Although this aspect will not be discussed here, it is certainly not forgotten.

Here, I have tried to come up with a use case that lies outside of the more obvious ones and that I personally would find interesting to investigate further.

Infant interaction design

What about interaction design for people who have yet to develop fine motor skills and speech?

Eye-tracking devices for infants[13] are becoming more common now but as far as I know it is for research applications and not for enabling infants to interact with electronic devices for their own sake.

[13] Children less than 12 months old.

While neonates[14] can sometimes track objects if they are large and move slowly, they seem to rely on saccades[15] rather than smooth pursuits. From about six weeks, smooth pursuits improves and infants begin to track things moving in a smooth pattern. The ability to track something moving in, say, a triangular wave comes later as that kind of sudden changes in direction demands better predictive notions (having an idea of where an object will move to next improves ones tracking capabilites).

[14] Children less than four weeks old.

[15] Rapid jumps between different fixation points in the same direction, usually occuring several times per second. This is a different type of eye movement than smooth pursuits.

At five months, smooth pursuits reaches adult performance when it comes to sinusoidally and horizontally moving targets.

> [G]aze and SP in the 1D horizontal condition in Study I is mature from the earliest age tested (5 months). The eye is on the object and the infants are able to predict the upcoming motion.[16]

[16] Grönqvist, 2010, p. 53.

Vertical tracking ability takes longer to develop and only at about 9 months is it comparable to the horizontal ability (although even in adults, the horizontal ability continues, in some regards, to be superior to the vertical)[17]. If horizontal and vertical movement is combined, like when tracking a circular motion, the vertical ability seems to be more negatively impacted compared to the horizontal[18].

[17] Ibid., pp. 36–37.

[18] Ibid., p. 52.

All in all, Pursuits seem to be a potentially good match with infants, especially if limited to smooth horizontal movement patterns. As their motor skills develop slower, there should be a window for gaze-only interaction design of *at least* six months. What is especially interesting is that this seems to be uncharted territory:

> The existing literature suggests that young children may benefit from using computers under the right circumstances. At the same time, it is clear that we know very little about the use of computers by children under the age of three, and even less for children under the age of two.[19]

[19] Hourcade et al., 2015, p. 1917.

Hourcade et al., 2015 analyzed videos on YouTube to investigate with what ability infants and toddlers[20] used tablets. It turns out that while most children under 12 months did not seem able to make meaningful use of the devices, the majority of children 12–17 months had reached moderate ability, and by the age of two, 90% had (note figure 3).

[20] Children one to three years of age.

Figure 3: The author added to the available research corpus by uploading a video of Harry, 6.5 months old, playing a game on the iPad 10.5" (Nockert, 2017).

Now, infants facing difficulties using touch-based devices is consistent with their development of fine motor skills. The study does however give some indication as to their desire to use electronic devices, were they able to.

> [T]here is a clear opportunity to explore the design of apps for very young children, starting as early as 12 to 17 months of age. The evidence presented in this paper suggests a majority [of] children in this age group and older can understand and use basic apps. The research question is how to design them such that they have similar characteristics to beneficial television shows, while helping children build communication, visual, and motor skills, and increase their connections to their caregivers.[21]

[21] Hourcade et al., 2015, p. 1923.

For me, there are a few particularly interesting aspects. First; if we want to know if it is possible for infants to learn to use applications equipped with gaze interaction, what kind of interfaces do we construct in order to try to find that out?

Second; if successful, what do we do with that knowledge? What interesting research questions could we then hope to answer?

Third; what if we could get these gaze-driven applications in the hands of (millions of) parents through tablets, mobiles and the web? It is here that I think Pursuits could be extremely useful. The built-in video capabilites of most devices today should be good enough to identify gaze-following (at least with exaggerated motions), and since there is no calibration step to go through (which I imagine would be difficult when it comes to infants), the big technical obstacles may very well be out of the way.

The thing is that it would really only take one child to provide a starting point for iterative improvements. Potentially opening up a branch of user interaction and experience design, one that has to be built from a new set of principles.

Looking at the iOS App Store, there are already many apps for use by infants (visual development, languages, mathematics, etc.)

but none that put them in control. Of course, this could very well be because infants lack the necessary drive or cognitive functions at this stage but given the sources referenced within, it seems more likely due to lack of an appropriate interaction technique.

Pursuits and Orbits as UI

Relying on moving controls may not be the most intuitive interface for adults used to conventional interfaces but for infants, it has advantages. To start with, once infants start developing smooth pursuits, they will also try to follow movement. Secondly, if they succeed to select a control, we know it is the result of an active eye movement rather than the child just happening to stare at something for a long enough time, accidentally triggering a gaze target.

One could also hope that it provides a clear causality — *I looked at the moving thing and then an interesting thing happened to it*. Research does point to infants having a sense of causality: "6-month-old infants can recognize the presence of causality embedded in a direct launching event"[22] and "results support the hypothesis that the human system possesses an early available, possibly innate basic mechanism to compute causality"[23], among others. In this perspective, it is not inconceivable that an infant actually could learn to interact with an application through gaze.

[22] Bélanger and Desrochers, 2001, p. 11.

[23] Mascalzoni et al., 2013, p. 327.

In one of the papers on Pursuits[24], a waiting room game is described, where a player helps a frog catch flies by gaze-following their flight paths. While sinusoidal elements provide advantages in moving smooth and as expected — we are free to use other paths as well. Any kind of moving object can be used as an interactive control as long as it moves in a way that we can match with corresponding eye movements. I think this provides us with the clearest vision of gaze-based UIs for small children to come.

[24] Vidal et al., 2013, p. 444.

Ethics

Children's use of electronic devices and apps for children, in general, are much debated, and especially so commercial apps. If we would be able to construct an interface that lets infants use electronic devices on their own, that is something which can not be put back in the box again. On the one hand, we should think long and hard about the consequences. On the other hand, if it is possible, it is only a question of *when* it happens. Personally, I think the possible benefits in terms of development and learning seem by far to outweigh the potential problems.

References

Bélanger, N. D. and S. Desrochers (2001). "Can 6-month-old infants process causality in different types of causal events?" In: *British Journal of Developmental Psychology* 19.1, pp. 11–21. ISSN: 2044-835X.

Esteves, A. et al. (2015a). "Orbits: Enabling Gaze Interaction in Smart Watches Using Moving Targets." In: *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. UbiComp/ISWC'15 Adjunct. Osaka, Japan: ACM, pp. 419–422. ISBN: 978-1-4503-3575-1.

– (Sept. 2015b). *Orbits: Gaze Interaction for Smart Watches*. Youtube. URL: https://www.youtube.com/watch?v=0xDAmOecFIU.

– (2015c). "Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements." In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Daegu, Kyungpook, Republic of Korea: ACM, pp. 457–466. ISBN: 978-1-4503-3779-3.

– (Nov. 2015d). *Orbits: Gaze Interaction for Smart Watches using Smooth Pursuit Eye Movements*. Youtube. URL: https://www.youtube.com/watch?v=xecOEK-kxMU.

Grönqvist, H. (2010). "Visual motor development in full term and preterm infants." PhD thesis. Uppsala University. ISBN: 978-91-554-7892-6.

Hourcade, J. P. et al. (2015). "Look, My Baby Is Using an iPad! An Analysis of YouTube Videos of Infants and Toddlers Using Tablets." In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, pp. 1915–1924. ISBN: 978-1-4503-3145-6.

Mascalzoni, E. et al. (2013). "The cradle of causal reasoning: newborns' preference for physical causality." In: *Developmental Science* 16.3, pp. 327–335. ISSN: 1467-7687.

Nockert, J. (June 2017). *Harry, 6.5 months old, using the iPad 10.5"*. Youtube. URL: https://www.youtube.com/watch?v=2ee0sW2_OC0.

Pfeuffer, K. et al. (2013). "Pursuit Calibration: Making Gaze Calibration Less Tedious and More Flexible." In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST '13. St. Andrews, Scotland, United Kingdom: ACM, pp. 261–270. ISBN: 978-1-4503-2268-3.

Tobii Pro X3-120 Accuracy and precision Test report (Sept. 2015). Tobii. URL: http://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-x3-120-accuracy-and-precision-test-report.pdf/?v=1.0.

Vidal, M. (Mar. 2013). *Pursuits: Spontaneous Interaction with Displays*. Youtube. URL: https://www.youtube.com/watch?v=TTVMB59KvGA.

Vidal, M. et al. (2013). "Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets." In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '13. Zurich, Switzerland: ACM, pp. 439–448. ISBN: 978-1-4503-1770-2.

– (Jan. 2015). "Pursuits: Spontaneous Eye-Based Interaction for Dynamic Interfaces." In: *GetMobile: Mobile Comp. and Comm.* 18.4, pp. 8–10. ISSN: 2375-0529.