Using the User's Point of View for Interaction on Mobile Devices

Jérémie Francone, Laurence Nigay

To cite this version:

Jérémie Francone, Laurence Nigay. Using the User's Point of View for Interaction on Mobile Devices. IHM 2011 - Conférence Francophone sur l'Interaction Homme-Machine, Oct 2011, Nice - Sophia Antipolis, France. pp.25-31, 10.1145/2044354.2044360. hal-00760343

HAL Id: hal-00760343
https://hal.science/hal-00760343

Submitted on 4 Dec 2012

Using the User's Point of View for Interaction on Mobile Devices

Jérémie Francone, Laurence Nigay
Université Joseph Fourier Grenoble 1
Laboratoire d'Informatique de Grenoble LIG UMR 5217, Grenoble, F-38041, France
{jeremie.francone, laurence.nigay}@imag.fr

RESUME

Nous présentons une technique d'interaction pour dispositifs mobiles (smartphone et tablette) basée sur le suivi du visage de l'utilisateur. Cette technique définit de nouvelles possibilités pour l'interaction en entrée et en sortie sur dispositifs mobiles. En sortie, le suivi de la tête peut permettre de contrôler le point de vue sur une scène 3D affichée à l'écran (Head-Coupled Perspective, HCP). Cette technique améliore l'interaction en sortie en offrant la perception de la profondeur et en permettant la visualisation d'un espace de travail plus grand (fenêtre virtuelle). En entrée, le suivi des mouvements de la tête définit une nouvelle modalité d'interaction qui ne requiert pas d'autres capteurs que la caméra du téléphone ou de la tablette. Dans cet article, nous explicitons les possibilités interactionnelles offertes par le suivi de la tête de l'utilisateur sur téléphones ou tablettes, particulièrement adapté au caractère mobile des dispositifs visés. Nous focalisons ensuite sur l'interaction en sortie en présentant plusieurs applications du HCP et en décrivant les résultats d'une expérimentation qualitative sur téléphone et tablette.

Mots clés

Dispositif Mobile, Modalité d'interaction, Interface 3D, Head-Coupled Perspective.

ABSTRACT

We study interaction modalities for mobile devices (smartphones and tablets) that rely on a camera-based head tracking. This technique defines new possibilities for input and output interaction. For output, by computing the position of the device according to the user's head, it is for example possible to realistically control the viewpoint on a 3D scene (Head-Coupled Perspective, HCP). This technique improves the output interaction bandwidth by enhancing the depth perception and by allowing the visualization of large workspaces (virtual window). For input, head movement can be used as a means of interacting with a mobile device. Moreover such an input modality does not require any additional sensor except the built-in front-facing camera. In this paper, we classify the interaction possibilities offered by head tracking on smartphones and tablets. We then focus on the output interaction by introducing several applications of HCP on both smartphones and tablets and by presenting the results of a qualitative user experiment.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**] User Interfaces – *Interaction styles*. I.3.6 [**Computer Graphics**] Methodology and Techniques – *Interaction techniques*.

General Terms

Design, Human Factors.

Keywords

Mobile Device, Interaction Modality, 3D Interface, Head-Coupled Perspective.

1. INTRODUCTION

Mobile devices such as smartphones and tablets are multifunctional. They enable us to run complex applications including video editors, web browsers and 3D games. As a consequence, more and more commands and data have to be displayed and managed, nearly reaching the processing capacity of a laptop, while the interaction capacity (input/output modalities) is limited due to the small form factor and the various contexts of use (e.g., walking in the street being encumbered by bags). Indeed, the particularity of these devices, that they have to be mobile, implies a limited screen size, no mouse and a lack of physical buttons. Moreover, as the user is potentially in a mobile situation, s/he is not likely to use both hands [18] and can devote only a limited attention to interacting with the application.

But as opposed to a laptop, being mobile means that the user can move and the device can be moved, which should be exploited for interaction. Current mobile devices have built-in multi-axis accelerometers, gyroscopes and sometimes a digital compass. These motion sensors are often used for sensing the screen orientation, controlling a character in a game or orienting a map. Since the seminal work of Rekimoto [25], there have been many systems that are based on motion sensors for enriching the interaction capacity of mobile devices. Amongst them, one relevant way to take advantage of the mobility is to use spatially aware displays [13] [14], which transform a position-tracked device into a window on a larger virtual workspace. By moving the device around, the user can access different parts of the workspace.

In this paper, we study the use of the spatial relation between the head and the device. In this context, the front-facing camera of the device is used to localize the face with respect to the device. By using vision-based algorithms, we have implemented an efficient

IHM'11, October 24-27, 2011, Sophia Antipolis, France

face tracker that improves the interaction for both input and output modalities. For input, we structure our study according to the characteristics of a modality based on the Theory of Modality [4] and present several examples of applications. For output, we focus on a technique called Head-Coupled Perspective that emphasizes relevant properties for mobile devices and that we have implemented and evaluated.

The paper is organized as follows. In the first part, we present our face tracking system. Then we explain the interaction possibilities offered by this technique on mobile devices for input and output.

2. FACE TRACKING

We have designed our technique to be used on existing mobile devices with no extra accessory. It works in real time on the iPhone 4 and the iPad 2 using their built-in front-facing camera, in both portrait and landscape modes.

Our face tracking system has been developed with the Open Computer Vision Library (OpenCV) [22], a cross-platform library mainly dedicated to real-time computer vision. We compiled it for iOS 4 devices. Our tracking system combines a Haar face detection technique and an extended Camshift tracking algorithm to perform a real time face tracking. Whenever a face is detected, the tracking starts and runs in background, allowing enough computational resources for the main tasks. The tracker generates the 3D face position as an output. The 2D position is computed from the location of the bounded box of the face in the picture with an acceptable accuracy and stability. In this way, whenever the device and/or the user's face actually move, the output value always matches the relative position of the user's face with regards to the device. For the moment, the distance from the face to the screen is naively estimated from the size of the bounded box, which is not working well as soon as the face does not fit entirely in the picture (Figure 1.c). As a consequence, the z-position of the virtual camera has been set to approximately thirty centimeters in our demonstrations during the experimental evaluations. Further ameliorations will try to improve this feature by seeking characteristic points on the face.

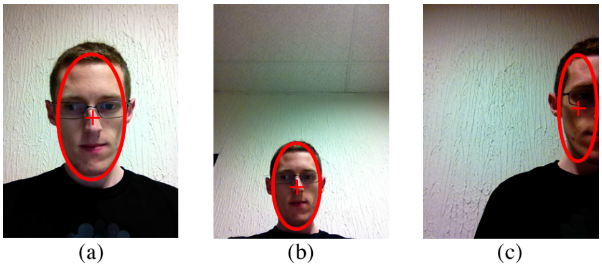
(a) (b) (c)

Figure 1. (a) (b) Examples of the face-tracking feature. (c) The face does not entirely fit into the picture, which is an issue for computing the distance of the user's head from the screen.

We make the assumption that there is one and only one face in the field of vision of the camera (Figure 1), which is likely to be the case on a mobile device. However, the tracker has to deal with the limited field of view and focal length of the built-in camera. If the user's face goes out of shot, we pause the tracking until the face comes back in shot. If the user still looks at the display while being out of shot, s/he may for example observe an incorrect perspective effect on screen. In addition, as we focus on the interaction, we do not specifically look for a robust tracking under extreme or varying lighting conditions. Some clues on how to implement a robust face tracker can be found in [8], [14] and [27].

3. INPUT MODALITIES

Face-tracking techniques enable us to create a spatially aware display, locating the device with respect to the face of the user. It is possible to take advantage of this relative position (user's head, device) for input interaction. Indeed face tracking provides information on the user's point of view on the display and also on the relative moves of the device according to the user's head.

To characterize the possibilities offered by face tracking for input interaction, we employ the two levels of abstraction of an interaction modality as defined in [21]: an interaction modality is defined as the coupling of a device d with an interaction language l: (d, l). The device d corresponds to the face tracking mechanism providing the 3D face position as an output for the input interaction language l. We study the characteristics of the interaction language l based on the Theory of Modality [4]. Indeed this theory defines a set of characteristics of output modalities and we apply them to the case of input modalities as done in [5]. We consider three characteristics of the input modality:

- *Arbitrary* -the modality does not rely on an already existing system of meaning-, or *non-arbitrary*.
- *Analogue* -there is a relation of resemblance with reality, for example a real-world metaphor-, or *non-analogue*.
- *Linguistic* -there is a structured system of signs that have a function of communication-, or *non-linguistic*.

These three characteristics allow us to classify the resulting gestural input interaction modalities that rely on face tracking.

Defining arbitrary modalities, several studies focused on the use of the <face-device> spatial relation for input interaction. The face location for a pointing task has been studied, in particular for selection in an egocentric circular menu [6] or for moving a cursor [10]. Obtained with a set of sensors fixed on the user's head (Shake device), the inclination of the head is used for interaction in a mobile situation [10]. In these studies, the interaction language of the gestural modality based on face tracking or on other sensors is arbitrary, non-analogue and non-linguistic.

The Perceptual Window [3] uses head tracking as a second input device in addition to the mouse on a desktop, which improves the performances for common GUI tasks. It consists of a natural gesture for reading a document (slight upward or downward moves in order to scroll). In the same way, Dynamic Peephole Display [33] uses various sensors to estimate the relative position of the device in relation to the user. It allows the user to navigate through a large information space by moving the device around. In particular, the modality enables the user to control the zoom factor by moving the device away/closer. As the gestures seem as natural as gestures done in the real world (e.g., approaching a document for reading it), the resulting modalities are analogue and hence non-arbitrary.

Finally, a linguistic modality such as "nodding head" (lower, raise head to node an assent) has been studied for hands-free interaction with a mobile device [20].

While the spatial relation between the user and the device has been studied for defining arbitrary, analogue or linguistic input modalities, we observe that on mobile devices, the modalities are not based on face tracking. Our work opens a vast new set of possibilities for input modalities based on face tracking, that we are currently exploring. As for inputs, a few studies exist on output modalities based on face tracking on mobile devices. This observation has motivated us to also study output modalities by focusing on interactive 3D views.

4. OUTPUT MODALITIES

By automatically altering the display according to the relative position of the device with the face of the user, we provide a natural and interactive 3D view: the user has the illusion to look at a small window instead of a flat screen. This technique is called Head-Coupled Perspective (HCP). From the user's point of view, the displayed objects are no longer flattened on the screen, but can stick out in front of the screen or be placed behind the screen surface. By slightly moving around her/his head and/or by tilting the device with her/his hand, the user can therefore estimate the depth position and the thickness of the displayed objects. S/he can also reveal objects hidden behind another one or behind the borders of the screen. Intuitively, this head-coupled perspective system fits well with mobile interaction and with 3D user interfaces. Moreover combining a 3D interface with a system that automatically maps the view perspective with the head's position enriches both the input and the output interaction bandwidths.

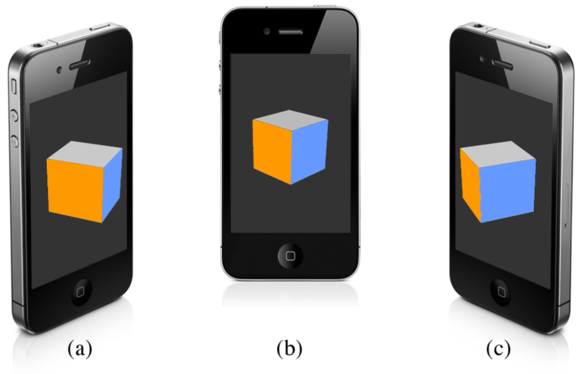

Figure 2. The cube is displayed at the screen level. The front part of the cube is therefore displayed at the front of the screen. If the user looks at the cube from the left (a) or from the right (c), s/he has the illusion that the cube pops out of the screen.

4.1 Related work

A review of the literature shows that many interaction techniques that rely on HCP have been studied since 1993. Head-Coupled Perspective view, also called Fish Tank Virtual Reality (VR) [32][1][24], consists of tracking the user's head position relative to the display in order to dynamically update a 3D projection matrix. With such vision-based HCP techniques, the user looks through the screen as if looking into a fish tank. This technique has been used for many years especially in the field of Virtual Environments to experiment on large and immersive displays. CAVE [11] and CABIN [17] are two examples. HCP has also been studied for desktop video games to help players do more tasks simultaneously and for improving attractiveness [29][30]. Yim et al. [34] proposed a head tracking game built upon the popular work of Lee [19].

On mobile devices, most research focuses on defining efficient and robust algorithms [8] [14] for head tracking (with varying lighting conditions and low computational cost), but only a few use it for HCP. pCubee [31] is a device that arranges five small screens into a small box shape and combines it with HCP technique. An experiment has shown promising results for visualization as well as for interaction in terms of performances and satisfaction. On a regular (one display) camera-equipped mobile device, face-tracking techniques have not been used for HCP. We note however a Nintendo DSi Ware game based on head-coupled perspective that has been released in Japan in 2010 [16]. In summary, there has been little or no work on Head-Coupled Perspective view for smartphones and tablets.

4.2 Rational for the design of our technique

Facing the ever-growing size of information spaces to be displayed on handheld devices, the screen size cannot be raised extensively due to mobility constraints, which leaves little space for the visualization of large information spaces. On the more recent devices, the pixel density has been extended to more than 300 pixel-per-inch (ppi), which approaches the resolution of the human retina at a reasonable distance. To sidestep these physical limitations, one solution is to enrich the visual language l of the output modality defined as (*screen*, l) by improving the way the information is encoded.

Our interaction technique adapts the display according to the way the user looks at the screen, creating a realistic perspective effect. It enables us to create a "virtual window": looking at the screen from the right, displays what is on the left by transforming the perspective, which allows us to observe a larger workspace. It is therefore relevant to combine this system with 3D user interfaces. To observe the 3D effect, the user must observe the scene from different points of view, which is coherent with the mobility constraints. The technique uses the head position to dynamically update a 3D projection matrix. By providing a natural interaction metaphor, our technique enriches the output interaction bandwidth. As opposed to auto-stereoscopic techniques where the user should stay still in front of the screen, the user can freely move. It is a monocular 3D display, each eye receiving the same picture.

With our technique, we increase the output interactional bandwidth by allowing a better depth perception and by enabling us to use 3D user interfaces efficiently. In this context, our technique improves the output bandwidth by using the depth as an extra dimension for encoding information. Indeed, our technique creates several new ways for encoding the information: for instance, giving one object a thickness, positioning a hierarchy of items in terms of depth, displaying stacks of objects, bringing out an important message in front of the interface as well as using shadows, occlusions and perspective in the display. Moreover Rekimoto [24] has shown that such a head-coupled perspective technique can improve the human's ability to understand complex 3D structures. It has been shown [12] that allowing the users to retain a control over the navigation helps them to familiarize with the 3D world.

In order to rationalize the design of our technique from the information visualization point of view, we characterize our technique in terms of operators with respect to the Visualization Framework of Ed Chi [9]. This framework is an operator and user-interaction model based on the visualization pipeline. Using this framework enables us to accurately characterize our technique and provides important clues on how to apply and implement it.

We highlight two high level operators that characterize our technique. Both of them are located at the end of the visualization pipeline and therefore tend to be domain/data independent:

- Perspective Transform operator: This operator is responsible for transforming the view perspective according to the face position on the current camera's frame. It is operationally similar across applications, which means that its implementation remains the same from application to application. It is also view-oriented, as it does not modify the data set. With respect to the Visualization Framework of Ed Chi, this operator takes place at the "View Stage" of the visualization process. It is therefore applicable to a wide range of application domains.
- 3D operator: This coarse-grain operator is in charge of the 3D scene and of the visualization of items. As this generic operator is able to operate on multiple data types, it should be further defined by multiple descriptions and consequently implementations. This operator is view-oriented and takes place at the "Visual Mapping Transformation Stage" of the visualization pipeline. By being closely related to the view, this operator encourages direct manipulation, which is a benefit for interactivity.

These visualization operators are very close to the "View Stage" (the last step in the visualization process). Therefore, they can be applied to many types of data. In the following section, we classify the possible applications of our technique with respect to different types of data.

4.3 Classification of the applications of our technique

We identify two orthogonal dimensions for classifying the use of our technique in interactive systems on hand-held devices (smartphones and tablets): the first one describes the use of the z-axis for displaying items; the second one describes the use of the z-axis for visualizing a workspace that is larger than the screen frame. Along each dimension, we classify the applications of our technique according to the types of data and tasks, as identified by "the task by data type" taxonomy of Shneiderman [28]. This taxonomy is based on seven data types (one-, two-, three-dimensional data, temporal and multi-dimensional data, tree and network data) and seven tasks (overview, zoom, filter, details on-demand, relate information, history of performed actions and extract part of the information).

4.3.1 Use of the z-axis for displaying items

The relevance of visualizing intrinsically three-dimensional objects with head-coupled perspective (HCP) technique is obvious. As explained in [24], being able to look at a 3D object from different points of view eases the understanding of the 3D structure. On a mobile device, 3D objects could be used in 3D-maps, games and scientific applications. Depending on the z-position of the object, it can be set at screen level, behind the screen, or even in front of the screen. Figure 2 shows a simple cube being displayed at screen level: the front part of the object is therefore displayed at the front of the screen surface, which gives the user the illusion that the cube pops out of the screen.

HCP technique is also relevant for non three-dimensional data. In this case, the depth is used as another means to encode the information. Since the head-coupled perspective technique drastically improves the depth perception, it is possible to express relationships among items (task: relate information) by varying their z-position. In this way, hierarchies of objects could be visualized. Moreover by providing a thickness to an object, natural metaphors can be applied. For example, the thickness of an email icon could increase as new messages are received. Figures 3 and 4 show an example application that uses the thickness of the items to express the ranking of each player with a podium metaphor. Another possibility is to hide one object behind another one: the hidden object is revealed if the user looks at the screen from the right point of view. Figure 5.c shows an example of this possibility.


(a) (b) (c)

Figure 3. An ordered list of player icons is displayed (screenshots of the iPad version (a) (b) and iPhone version (c)). Each item has a thickness that indicates the ranking of the player with a podium metaphor. (a) The user looks from the front; (b) the user looks from the bottom; (c) the user looks from the top.

Figure 4. A non-ordered list of icons is displayed from different points of view with thickness as a means to express the ranking of the players (pictures of the iPad version).

4.3.2 Use of the z-axis for visualizing a large workspace

HCP technique not only enables us to use the z-axis as a new means to encode the visual information, it also enables us to visualize a workspace that is larger than the display. In fact, HCP creates a virtual window, at which it is possible to look, as with any actual window: looking from one side reveals what is on the opposite side. The only difference with an actual window is that it fits in the hands of the user. Looking at the display from different points enables us to reveal information that is virtually hidden around the display. For example, this could be combined with Halo [2] on a 2D map: the points of interest could be indicated only if the user looks in their direction, which avoids overloading the main part of the display. Another example consists of transforming the display into a virtual box to increase the informational density: it creates a kind of interactive fish-eye effect by displaying data on each side of the box. Figure 5 shows an example of this box metaphor: the rear side of the box displays the current page of icons while the right and left sides of the box display an overview of the previous and next pages; the front face is used for emphasizing important items.

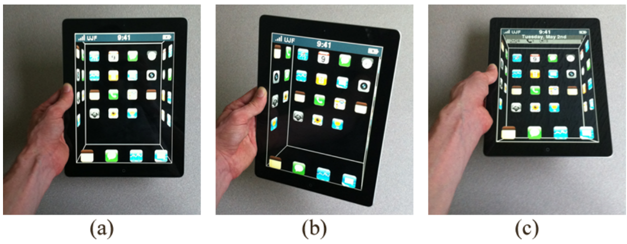

Figure 5. A grid of icons with a "box metaphor". By moving the device and/or her/his head, the user can (b) catch a glimpse of the neighboring pages and (c) reveal hidden items.

4.4 Implementation

We use OpenGL ES 2.0 [23] for the rendering. With traditional 3D graphics, the standard projection matrix assumes that the user is positioned directly in front of the display at a predefined distance. Conversely, Head-Coupled Perspective techniques use a custom projection matrix that is computed at each frame depending on the face position. This involves using a non-symmetric view frustum that provides the correct projection according to the point of view of the user. Such projections are called off-axis projections. In practical terms, we set the origin of the 3D scene at the center of the screen and we consider the screen frame's dimensions as constant values. In this coordinate system, a virtual camera is set at the position of the user's head, pointing towards the origin. As the user's head moves, the parallax of the 3D scene is altered by varying the virtual camera's field of view and its offset position (Figure 6). The user has then the illusion of handling a small window portrayed onto the 3D scene.

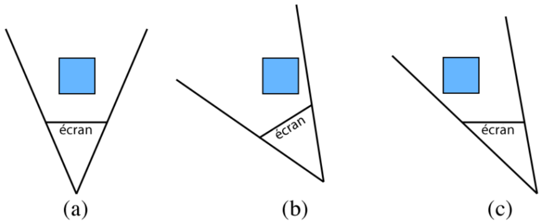

Figure 6. Comparison of standard display and HCP: (a) The user looks at the display from the front. (b) Looking from the right with a standard display only moves the projection volume. (c) Looking from the right with HCP distorts it (c).

4.5 Initial user evaluation

We conducted a qualitative user study with 20 unpaid participants ranging in age from 12 to 83. Half of the participants were recruited from within the university community while the others were not skilled at computer science (and in particular human-computer interaction domain). The goals of the experiment were to evaluate the usability and the user satisfaction of our technique on smartphones and tablets, and to collect a maximum of feedback.

A first experiment involving only the iPhone version demonstrated that the technique is usable and satisfying on smartphones. We have repeated the same protocol with both the iPhone and the iPad versions.

4.5.1 Experimental design

Participants were asked to use four example applications that illustrate the different possibilities of HCP. These four applications have been implemented in two slightly different versions, one being adapted to the iPhone's screen (3.5") and the other being adapted to the iPad's screen (9.7"). Half of the participants started with the iPhone version before evaluating the iPad version. The other half evaluated the iPad first.

The first application (Figure 7) was a simple 3D scene displaying several targets floating at different depth levels. Some targets were popping out of the screen while others were at screen level or behind the screen, as introduced by Lee [19] with its Wii Remote Head-Tracking demonstration. The goal was to make sure the participants understood the technique. After a short practice time, we asked the participant to describe what s/he was seeing.

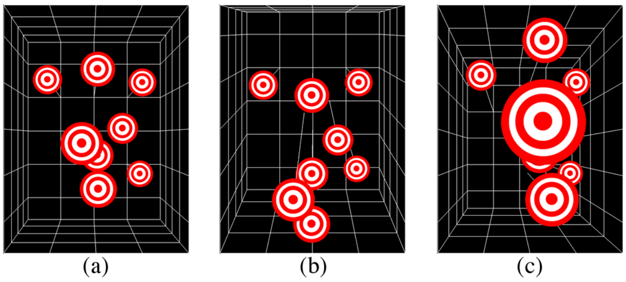

Figure 7. (a), (b) et (c) show three points of view of the same 3D scene. (c) The head-screen distance is taken into account and the face is close to the screen. This parameter has not been enabled during the experiment.

The second application was comprised of two parts: first, a simple 2D ordered grid of icons of players and their ranking position in a game was shown to the participants. Then, the same ordered grid of players was shown in 3D (Figure 3), each icon having a thickness that depends on its ranking using a podium metaphor: the first player was the thickest icon while the last player was the thinnest.

The third application was based on the same principle but the items were not ordered, making it more difficult to evaluate the relative depths (Figure 4).

The last application was a 3D version of the iPhone Springboard, displaying the apps icons using a box metaphor. Such a representation allows previsualization of the neighboring pages of icons and permits us to hide peripheral information behind the sides of the screen, which require the participants to look at the interface from different points of view. Figure 5 illustrates several points of view of this application.

Post-study written and verbal questionnaires were completed after each stage (first device, second device). The written questionnaires were based on System Usability Scale (SUS) [7] and AttrakDiff [15]. The users were encouraged to think aloud during the experiment. At the end of the experiment, participants were asked to fill one last questionnaire for comparing the smartphone and the tablet versions. We also asked participants to give three good points and three bad points about the technique.

4.5.2 Results and discussion

SUS scores are similar for both the iPhone and the iPad versions: 83.4 (4th quartile) in terms of usability, which means the technique is acceptable and almost “excellent”. Attrakdiff measures the attractiveness of an interactive system. It provides detailed data on usability, hedonics quality and attractiveness. It enables us to compare several interactive systems. Results of Attracdiff show that there is no significant difference between both versions. They indicate that our technique is “rather desired” (Figure 8) with a very small confidence rectangle, meaning that the results are reliable and not coincidental. The technique is in the above average region for each dimension of Attrakdiff, which means it is an attractive technique that stimulates users and awakes curiosity. Similar results were shown during the previous experiment that involved only the iPhone version for SUS and Attrakdiff.

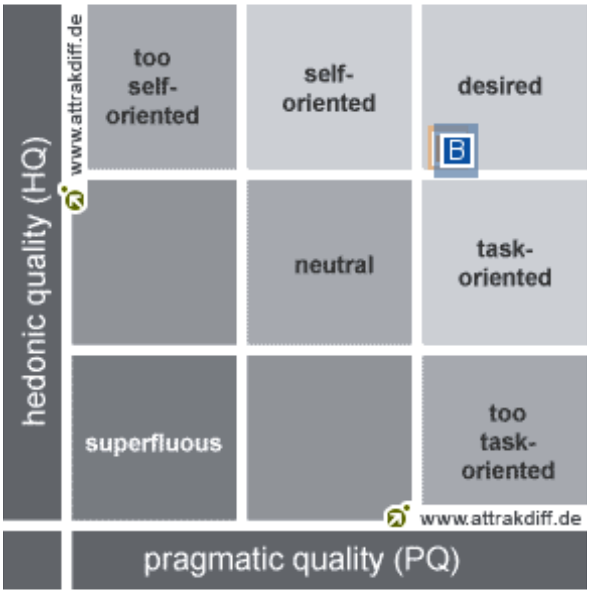

Figure 8. Diagram representing the main Attrakdiff results for our technique on iPhone and iPad. The two rectangles are bound up in the “desired” area.

14 participants said that the technique is natural and immersive. 6 participants were frustrated by the limited field of view of the built-in camera, which breaks the perspective effect at a certain angle if the face goes out of shot. It means that we should not design interfaces that require looking at the display from extreme angles. Another possibility could be to combine the HCP technique with additional sensors for trying to extend the effect, or to use an external camera with a wide-angle lens. Interestingly, almost every participant moved the device and did not move her/his head or body.

Every participant was able to see the depth effect, but only 9 have seen that some targets popped out of the screen. 12 participants have noticed that the tablet version provided a better 3D effect than the smartphone version. Every participant was able to indicate a ranking of the players in the second application, but they experienced some difficulties once the players were no longer ordered (third application). In this case, the participants were not always able to evaluate the difference in depth between two remote items. This means that we should perform quantitative experiments to precisely evaluate depth perception with our technique. About the last application, 6 participants said that the icons displayed on the rear face of the box were moving too fast. This issue has not been noticed on the tablet version because the icons are bigger. This means that small interactive objects should not be too far from the projection plane (i.e. from the screen surface).

In summary, this experiment showed that head-coupled perspective technique is an acceptable technique for smartphones (small screen, very mobile) as well as for tablets (bigger screen, less mobile). Participants were able to understand the depth effect, and perceived it as natural, attractive and innovative. However, there is room for improvement, especially for overcoming the limitations of the built-in camera and for improving the depth perception.

5. CONCLUSION AND FUTURE WORK

Finding new ways to improve the input and output interaction capabilities of mobile devices is an important challenge. In our study, we addressed this dual challenge by focusing on a non-intrusive face-tracking technique on smartphones and tablets. It is noticeable that the hardware required for this technique is already available on many mobile devices.

For input, we highlighted the vast set of possibilities by classifying the input modalities based on face tracking according to three characteristics (arbitrary, analogue and linguistic).

For output, we introduced a face-tracking head-coupled perspective technique that transforms the flat display of a smartphone or a tablet into an interactive monocular 3D display. With this system, the user perceives the 3D depth and has the illusion to look at a small window. We highlighted the great potential of the technique by classifying its different applications on mobile devices. Moreover the qualitative user study based on four example applications of the technique on both a smartphone and a tablet demonstrated the ease and naturalness of the technique. This makes it a promising interaction technique for mobile devices. An iOS application called “i3D” has been released on the App Store in order to demonstrate our technique.

Face tracking opens a vast set of possibilities for pure or combined input modalities: one important strand of our future work is to use our technique as an input modality for managing large information spaces in the context of a multimodal 3D interface. As a first attempt, we are currently studying the perspective wall [26] coupled with our technique on smartphones and tablets.

6. REFERENCES

[1] Arthur, K.W., Booth, K.S., Ware, C. 1993. Evaluating 3D Task Performance for Fish Tank Virtual Worlds. *ACM Trans. Inf. Syst.* 11, 3 (July 1993), 239-265. DOI= http://doi.acm.org/10.1145/159161.155359

[2] Baudish, P., Rosenholtz, R. 2003. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '03).* ACM, New York, NY, USA, 481–488. DOI= http://doi.acm.org/10.1145/642611.642695

[3] Bérard, F. 1999. The Perceptual Window: Head Motion as a New Input Stream. In *Proceedings of INTERACT 1999*, IOS Press, 238-244.

[4] Bernsen, N., 1994. Modality Theory in Support of MultiModal Interface Design. In *Proc. of Intelligent Multi-Media Multi-Modal Systems*, 37-44.

[5] Bouchet, J., Nigay, l., Ganille, T. 2004. ICARE Software Components for Rapidly Developing Multimodal Interfaces. In *Proceedings of ICMI 2004*. ACM, New York, NY, USA, 252-258. DOI= http://doi.acm.org/10.1145/1027933.1027975

[6] Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S. 2003. Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. In *Proceedings of CHI 2003*. ACM, New York, USA, ACM, 473-480. DOI= http://doi.acm.org/10.1145/642611.642694

[7] Brooke, J. 1996. SUS: A Quick and Dirty Usability Scale, in P.W. Jordan, B. Thomas, B.A. Weerdmeester & I.L. McClelland (Eds.), Usability Evaluation in Industry. London, UK: Taylor & Francis. 189-194.

[8] Bulbul, A., Cipiloglu, Z., Capin, T. 2009. A Face Tracking Algorithm for User Interaction in Mobile Devices. In *Proceedings of CW 2009*. IEEE Computer Society, Washington, DC, USA. 385-390. DOI= http://dx.doi.org/10.1109/CW.2009.9

[9] Chi, E., Riedl, J. 1998. An Operator Interaction Framework for Visualization Systems. In *Proceedings of InfoVis 1998*, IEEE Computer Society, 63-70. DOI= http://dx.doi.org/10.1109/INFVIS.1998.729560

[10] Crossan, A., McGill, M., Brewster, S., Murray-Smith, R. 2009. Head Tilting for Interaction in Mobile Contexts, In *Proceedings of MobileHCI '09.* ACM, New York, NY, USA. Article 6. DOI= http://doi.acm.org/10.1145/1613858.1613866

[11] Cruz-Neira, C., Sandin, J.D., DeFanti, T.A. 1993. Surround-Screen Projection Based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of SIGGRAPH 1993*. ACM, New York, NY, USA. 135-142. DOI= http://doi.acm.org/10.1145/166117.166134

[12] Elmqvist, N., Tudereanu, M.E., Tsigas, P. 2008. Evaluating Motion Constraints for 3D Wayfinding in Immersive and Desktop Virtual Environments. In *Proceeding of CHI 2008*. ACM, New York, NY, USA. 1769-1778. DOI= http://doi.acm.org/10.1145/1357054.1357330

[13] Fitzmaurice, G.W. 1993. Situated Information Spaces and Spatially Aware Palmtop Computers. *Communications of the ACM*. 36, 7 (July 1993). 38-49. DOI= http://doi.acm.org/10.1145/159544.159566

[14] Hannuksela, J., Sangi, P., Turtinen, M., Heikkilä, J. 2008. Face Tracking for Spatially Aware Mobile User Interfaces. In *Procedings of ICISP 2008*. Springer-Verlag, Berlin, Heidelberg. 405-412. DOI= http://dx.doi.org/10.1007/978-3-540-69905-7_46

[15] Hassenzahl, M., Burmester, M., Koller, F. 2003. AttrakDiff: A Questionnaire for Measuring the Perceived Hedonic and Pragmatic Quality. *Human & Computer 2003*. 187-196.

[16] "Hidden 3D Image: There it is!", Nintendo DSi Ware, 2010. www.nintendo.co.jp/ds/dsiware/krgj/

[17] Hirose, M., Ogi, T., Ishiwata, S., Yamada, T. 1999. Development and Evaluation of the Immersive Multiscreen Display CABIN. *Systems and Computers in Japan*. 30: 13-22. DOI= http://dx.doi.org/10.1002/(SICI)1520-684X(199901)30:1<13::AID-SCJ2>3.0.CO;2-L

[18] Karlson, A., Bederson, B., Contreras-Vidal, J. 2008. Understanding One Handed Use of Mobile Devices. In *Lumsden, J. ed., Handbook of Research on User Interface Design and Evaluation for Mobile Technology*. IGI. 86-101.

[19] Lee, J.C. 2008. Hacking the Nintendo Wii Remote. In *IEEE Pervasive Computing* 7, 3 (July 2008). 39-45. DOI= http://dx.doi.org/10.1109/MPRV.2008.53

[20] Morency, L.P., Darrell, T. 2006. Head Gesture Recognition in Intelligent Interfaces: the Role of Context in Improving Recognition. In *Proceedings of IUI '06*. ACM, New York, NY, USA. 32-38. DOI= http://doi.acm.org/10.1145/1111449.1111464

[21] Nigay, L., Coutaz, J. 1995. A Generic Platform for Addressing the Multimodal Challenge. In *Proceedings of CHI '95*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA. 98-105. DOI= http://dx.doi.org/10.1145/223904.22391798-105.

[22] Open Computer Vision Library (Official website) http://opencvlibrary.sourceforge.net/

[23] OpenGL ES (Official website) http://www.khronos.org/opengles/

[24] Rekimoto, J. 1995. A Vision-Based Head Tracker for Fish Tank Virtual Reality - VR Without Head Gear. In *Proceedings of VRAIS '95*. IEEE Computer Society, Washington, DC, USA. 94-100. DOI= http://doi.ieeecomputersociety.org/10.1109/VRAIS.1995.512484

[25] Rekimoto, J. 1996. Tilting Operations for Small Screen Interfaces. In *Procedings of UIST '96*. ACM, New York, NY, USA. 167-168. DOI= http://doi.acm.org/10.1145/237091.237115

[26] Robertson, G.G., Mackinlay, J.D., Card, S.K. 1991. The Perspective Wall: Detail And Context Smoothly Integrated. *In Proceedings of CHI '91*. ACM, New York, NY, USA. 173-179. DOI= http://doi.acm.org/10.1145/108844.108870

[27] Ruddarraju, R., Haro, A., Nagel, K., Tran, Q.T., Essa, I.A., Abowd, G., Mynatt, E.D. 2003. Perceptual User Interfaces Using Vision-Based Eye Tracking. In *Proceedings of ICMI '03.* ACM, New York, NY, USA. 227-233. DOI= http://doi.acm.org/10.1145/958432.958475

[28] Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of VL '96*. IEEE Computer Society,

Washington, DC, USA. 336-343. DOI= http://doi.ieeecomputersociety.org/10.1109/VL.1996.545307

[29] Silva, M.G., Bowman, D.A. 2009. Body-Based Interaction for Desktop Games. In *Proceedings of CHI '09*. ACM, New York, NY, USA. 4249-4254. DOI= http://doi.acm.org/10.1145/1520340.1520648

[30] Sko, T., Gardner, H.J. 2009. Head Tracking in First-Person Games: Interaction Using a Web-Camera. In *Procedings of INTERACT '09*. Springer-Verlag, Berlin, Heidelberg. 342-355. DOI= http://dx.doi.org/10.1007/978-3-642-03655-2_38

[31] Stavness, I., Lam, B., Fels, S. 2010. pCubee: a Perspective-Corrected handheld cubic display. In *Proceedings of CHI '10*. ACM, New York, NY, USA. 1381-1390. DOI= http://doi.acm.org/10.1145/1753326.1753535

[32] Ware, C., Arthur, K., Booth, K.S. 1993. Fish Tank Virtual Reality. In *Proceedings of INTERACT '93 and CHI '93*. ACM, New York, USA. 37-42. DOI= http://doi.acm.org/10.1145/169059.169066

[33] Yee, K.P. 2003. Peephole Displays: Pen Interaction on Spatially Aware Handheld Computers. In *Proceedings of CHI '03*. ACM, New York, NY, YSA. 1-8. DOI= http://doi.acm.org/10.1145/642611.642613

[34] Yim, J., Qiu, E., Graham, T.C.N. 2008. Experience in the Design and Development of a Game Based on Head-Tracking Input. In *Proceedings of Future Play '08*. ACM, New York, NY, USA. 236-239. DOI= http://doi.acm.org/10.1145/1496984.1497033