AR Human Interfaces: The Case of Gesture

A position paper for the International AR Standards Meeting February 17-19, 2011

Christine Perey
PEREY Research & Consulting
cperey@perey.com
+41 79 436 68 69

***Abstract*—** Human interfaces with new technologies have grown more complicated and more difficult for mainstream users to adopt as the designers of new experiences seek to push ever more functionality and flexibility into smaller spaces.

Augmented Reality companies and groups studying gesture as a computer human interface should learn more about the challenges facing users who will use AR without keyboards or touch screens to control their experiences. We call on those in both communities to collaborate towards the development of intuitive and natural gestures permitting users to find, interact with and expand content in context.

I. INTRODUCTION

In his book, *The Design of Everyday Things* [1], Donald Norman points out that as technology advances the connections between the user's action and the results must be continually evaluated and it becomes increasingly difficult to balance simplicity and flexibility.

A "traditional" radio receiver has, at minimum, two dials: one for the volume and one for tuning. The user twists the dial to the right to go up and to the left to go down. If stereo, a third dial is frequently the interface to adjust the balance between right and left speakers. A twist to the left puts more sound in the left speaker and vice versa. Anyone learns how these dials work without instructions (universally) with the first experience because the feedback is immediate. With the introduction of pressure sensitive controls, users had to become accustomed to an alternative user interface: a button to the right caused the radio to detect stations broadcasting at higher band; pressing another button, to the left of the first, caused the radio to seek stations at a lower band. Some designers put the "up" button above the "down" button, but either horizontally or vertically arranged, the effect of the button was clear to the user by way of the display which shows the band selected rising or falling. And, regardless of the manufacturer, the user actions produce a familiar experience.

When the graphical user interface was first proposed as a way for users to interact with computers, user experience designers experimented with different metaphors. The concept of "windows" which could be overlaid and moved about, resized and opened/closed by dragging and "clicking", was an amalgam of many prior paradigms, while also introducing new ones. To a user who has never experienced a GUI, perhaps the "window" on a document was somewhat familiar because it mimics (to some extent) the use of sheets of paper. There are numerous examples of the interaction with, and the behaviors of objects or information in a digital environment imitating the real world.

There are also actions which have no physical world equivalent and which are not consistently implemented across all digital devices. Inconsistencies require a user to learn a new interface each time they change device. One of the biggest differences between the two dominant personal computer operating systems is the position of the "close, hide and resize" buttons. In the Apple operating system, these controls appear in the upper left corner of an active area; Windows applications use the upper right, a subtle but important distinction. One of the frequent reasons a consumer stays with one device manufacturer is they learn incrementally new skills/user interface options.

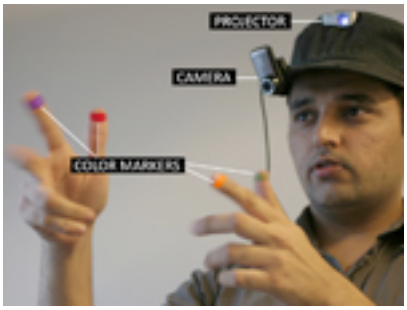
Just moving from a feature phone to a smartphone requires training. Norman refers to this as the curve of complexity. The flexibility/customizability of smartphones is deep, and the rewards high, if a user learns how to use them.

The purpose of this paper is to examine if and how AR human interfaces may become the **center of a larger collaboration across multiple domains** to lessen the steepness of the complexity curve when manufacturers introduce gestures as a mode of interaction with new devices, the real world and digital content.

II. INTERACTIVITY IN AR

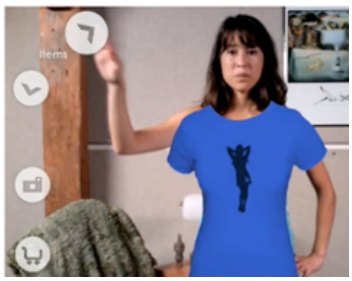
Today's consumer augmented reality interfaces rely heavily on the interaction paradigms of the platforms on which they run. Applications running on personal computers rely on the user holding something in front of the webcam; most AR on smartphones uses the touch interface to produce behaviors (e.g., touch once to select, touch twice to "open", pinch to shrink, etc). The user can also manipulate the position (pose) of the device relative to the real world, by turning (panning).

One solution illusrated by a project from Pattie Maes' Fluid Interfaces group [2], is to integrate off the shelf technologies such as a pico projector, a mirror, small webcam and the "mobile computing device" in the pocket. The Sixth Sense wearable gestural



interface project [3] caught the immagination of many people and the popular media, illustrating the appeal of gesture for controlling information search and discovery, as well as communication and other activities.

Some innovative desktop AR applications have also experimented with using gestures. For example, Zugara's AR dressing room application "Webcam Social Shopper" [2] permits the user to change colors and control the software features by moving arms and hands.



With the introduction of Microsoft Kinect, the laboratory experimentation with gestures has sky rocketed to new levels (too numerous to count or cite). Behavioral scientists and user experience designers are looking at gestures for AR as well as other device control functions.

III. User-Centered Design Principles

As is encouraged in the paper submitted for this meeting by Synthetic Toys [5], AR should be considered first and foremost a user experience. To drive the user experience development from a "purely" engineering-centric point of view (what can the technology do?) will handicap the field's expansion terribly. Likewise, to define AR experiences solely on what content is already available (images of advertisements, book jackets, CD and DVD covers in the case of vision-based AR and Lat/Long of landmarks and buildings in the case of geospatial AR) constrains the scope of this field unnecessarily.

There is no doubt that exploration and experimentation with Kinect will shed light on what constitutes good and bad gestures for use when the camera(s) and display are at a distance from the user. The introduction of eyewear for AR in the coming 12 to 18 month will fuel explorations of gestures when the camera and display are close to the user's head and AR experiences will begin to be "hands free." These two paradigms seem sufficiently different—the camera in the "distance" scenarios and use cases, and the "close up" scenarios in which primarily the user's hands and arms will be detected— to serve as the basis for developing strong principles in gesture.

Norman's Seven Principles of Transforming Difficult Tasks into Simple Ones is proposed as a framework for exploring the use of AR for introducing gesture into many facets of modern life. The following is an exercise to use these principles and apply them to the use of gestures with AR.

A. Use both knowledge in the world and in the head

It is difficult to apply this principle for using gestures with AR because users have relatively little knowledge of AR in the real world. But they *do* know how objects behave in the real world: an object rotates when a "spin" is applied with the twist of a hand; they move to the right, left, up or down when another object with greater force than gravity acts on them.

B. Simplify the structure of tasks

This has to be examined from three points of view:

1. The structure of the tasks available given the technological constraints
2. The structure of tasks available given the content available
3. The structure of tasks that the user might wish to achieve.

We suggest that certain tasks can be defined as universally desirable; for example: Search/find, Select/choose, Deselect/forget, Next/Previous, Annotate (with options such as: Favorite (or Hate), Collect (come back to later), Communicate (engage directly with the provider of the subject, send this information to third party).

C. Make things visible

This principle is quite natural for AR to implement since making the digital visible in the context of the real world is entirely its purpose.

D. Get the mappings right

Here Norman encourages adopting the response compatibility principle. "The major requirement of response compatibility is that the spatial relationship between the positioning of controls and the system or objects upon which they operate should be as direct as possible with the controls either on the objects themselves or arranged to have an analogical relationship to them." [6]

With real time natural feature detection and tracking technology, putting controls directly over or near the object of interest seems to be far more achievable than in 1988, (when *The Design of Everyday Things* was written).

What is more challenging is to apply this relationship to the system outputs. Here, user experience designers will need to provide visual or auditory confirmation when a gesture has been understood and there might be a delay between the request and the resulting appropriate action or behavior.

E. Exploit the power of constraints, both natural and artificial

Users are not expecting their applications to do everything for them all at once, at least not in the first year. Smartphones have already taught certain population segments that they must purchase and/or download applications in order to be able to achieve certain objectives with their technology platforms.

In this regard, having AR browsers that have few constraints, use "layers" or "channels" of information as the constraining module, this has worked quite well to date. What are the equivalent constraints with gestures?

For most vision-based features, the most significant constraint is that of lighting. In low light or very high light environments (producing shadows and high contrast), the technology constraints will be with users for many years unless complete solutions provide remedies. Here we could include, with the gesture hardware and software, colored finger caps, as used by Pranav Mistry in the Sixth Sense demonstration [3].

F. *Design for error*

Here, we must assume that there will be gestures that are incomplete, there will be occlusions and people with handicaps preventing their moving fingers in specific directions repeatedly. Norman encourages designers to plan for error as a natural, constructive dialog between the user and the system. "Allow the user to recover gracefully from errors. Make it easy to reverse operations; make it hard to do irreversible actions" suggests Norman.

G. *When all else fails, standardize*

On this point Norman is clear; when something cannot be designed without arbitrary mappings and difficulties, standardization is one solution to make experiences predictable. He reminds designers that the positions of the keys on keyboards, the numbers on the face of a clock and many other elements of everyday objects are standardized and have only to be learned once (for each language).

Standardization is another aspect of cultural constraints and can simplify lives as well as technology design. However, in this paper, we suggest that exploring gestures for AR is just in its very early phase. The field is far from reaching a point where standardization is recommended, but let us learn from those examples where standardization was too late, and less elegant solutions become widespread (e.g., the metric system is better than the British/US system of measure).

IV. RECOMMENDATIONS

It is time for professional AR experience designers to collaborate with the providers of content and with those in academia and industry with specific expertise in multimodal interfaces and gesture. These collaborations should develop studies using classical design research principles, a framework such as that proposed in another paper [5] and evaluate the most promising uses of gesture in AR.

More specifically, companies designing AR experiences should be sponsoring and receiving interns from academic groups, such as the MIT Fluid Interface Group, the Glasgow Interactive Systems Group [7], in particular the lab of Stephen Brewster, the A.R.A.M.I.S. (Augmented Reality Avatar as Model for Interaction Science) project [8] and other centers of gesture study around the world.

V. CONCLUSION

Gesture will be a popular method of interaction with digital content in the future, especially as the real world and digital content intersect. As content in context continues to merge in Augmented Reality experiences, there are many opportunities for new interactions to benefit from gesture as a method of input and user interactivity.

Within the AR community, many specialists are pushing technological envelopes and remain reluctant to engage with user behavior scientists and the owners of content to fully explore how gesture will be integrated in the future of AR. Such collaborations in our community are to be highly encouraged. We believe that when the teams focusing on gestures in AR are in place, the entire ecosystem, from the manufacturers of hardware to the providers of content will benefit.

ABOUT AUTHOR

Christine Perey has been working in the domain of rich media communications for 20 years, initially in the area of dynamic media technologies on personal computers; she founded, was the editor and publisher of the QuickTime Forum, a publication for QuickTime developers 1991-1993. When enterprise and wide area networks emerged as a means of distributing rich media created and consumed on personal computing devices, she expanded to provide a variety of services to the companies in the rich media communications industry, and accelerated adoption of videoconferencing and streaming media.

Since 2006, Perey studies and assists companies to better understand their opportunities in the domains of Mobile Social Networking and Augmented Reality. Perey is an active leader of new industry initiatives, organizing thought-leadership meetings and think tanks about mobile AR topics.

REFERENCES

[1] Norman, Donald A. (1988) *The Design of Everyday Things* Doubleday

[2] Fluid Interfaces Group http://fluid.media.mit.edu/

[3] Sixth Sense project
http://www.pranavmistry.com/projects/sixthsense/

[4] Fashionista project
http://www.zugara.com/projects/fashionista

[5] Standards for Augmented Reality: A User Experience Perspective by Panagiotis Ritsos, Dennis Ritsos, Antonis Gougoulis for International AR Standards Meeting Feb 17-19, 2011 http://www.perey.com/ARStandards/Ritsos-A_User_Experience_perspective.pdf

[6] Page 199, *The Design of Everyday Things*

[7] Glasgow Interactive Systems Group
http://www.dcs.gla.ac.uk/gist/

[8] A.R.A.M.I.S. Augmented Reality Avatar as Model for Interaction Science Project http://aramis.project.eia-fr.ch/