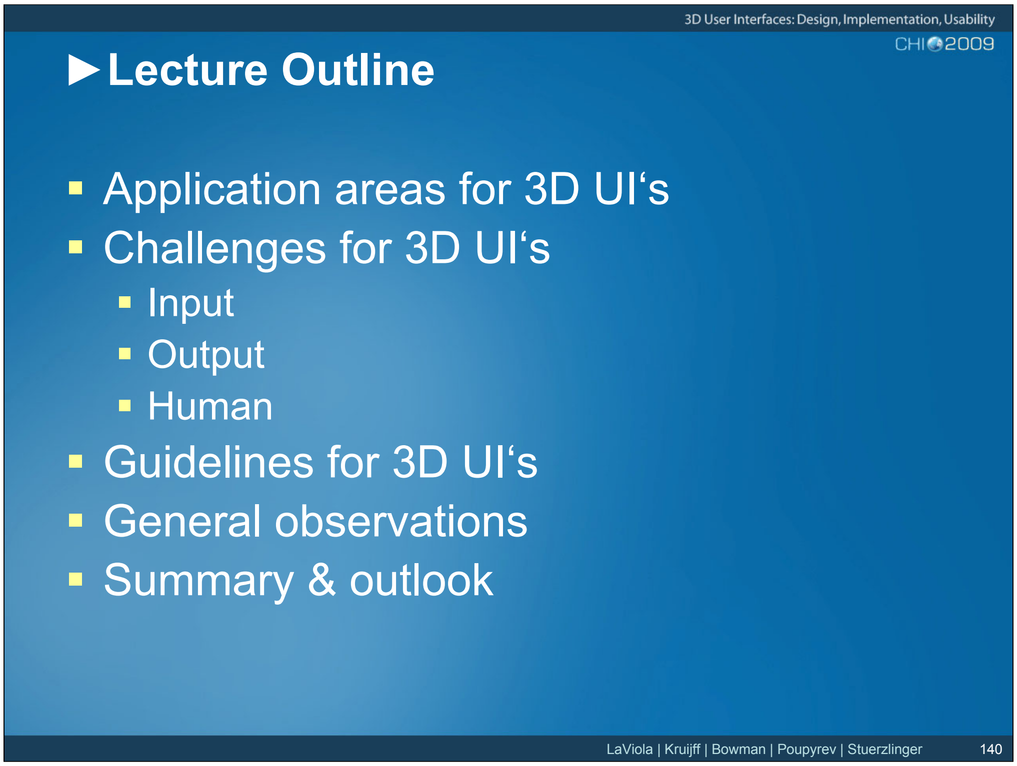
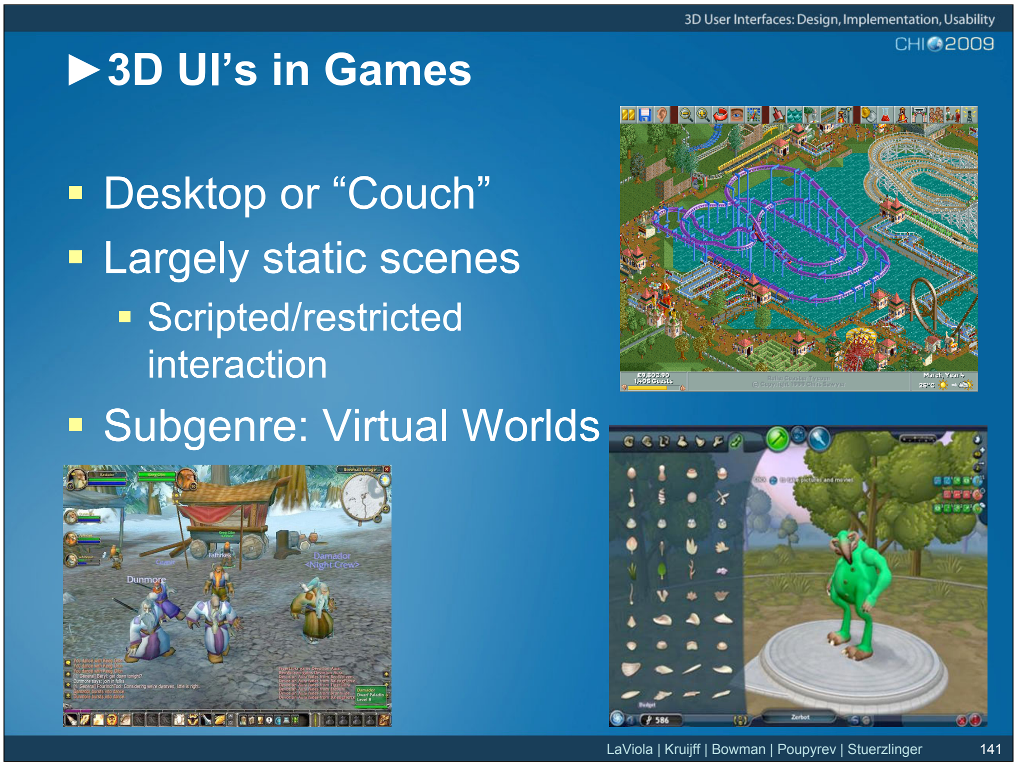
Guidelines for Developing 3D UIs

Wolfgang Stuerzlinger

Welcome, Introduction, & Roadmap
3D UIs 101
3D UIs 201
User Studies and 3D UIs
Guidelines for Developing 3D UIs
Video Games: 3D UIs for the Masses
The Wii Remote and You
3D UI and the Physical Environment
Beyond Visual: Shape, Haptics and Actuation in 3D UI
Conclusion



In this lecture, we will first briefly discuss the most important application areas of three-dimensional user interfaces (3D UI's) and review important characteristics of these areas. Then we look at technical issues around input and output hardware as well as important facts about humans that related to 3D UI's. Based on this reflection, we then derive a list of guidelines on how to make good 3D user interfaces. Finally, we look at the "big picture" as a guide design decisions.



3D Games:

-Goal: entertainment

-Two main settings:

    -Desktop setting

        -Keyboard & mouse

    -"Couch" setting

        -Relaxed sitting or standing

        -Game controllers

        -Recently, WiiMote & lot's of other external devices (guitars, WiiFit, etc.)

-Mostly static environments with a few "active" entities

-Often quasi-3D, i.e. greatly restricted environment

    -These restrictions used to great advantage, e.g. in Spore 3D creator

-Usually lot's of interaction, but lot's of restrictions (can't break through wall unless predefined)

Subgenre: Virtual Worlds

-Goal: social interaction, economy, etc.

    -Popularity has peaked to some degree

-Support for content creation, but typically very cumbersome 3D interfaces

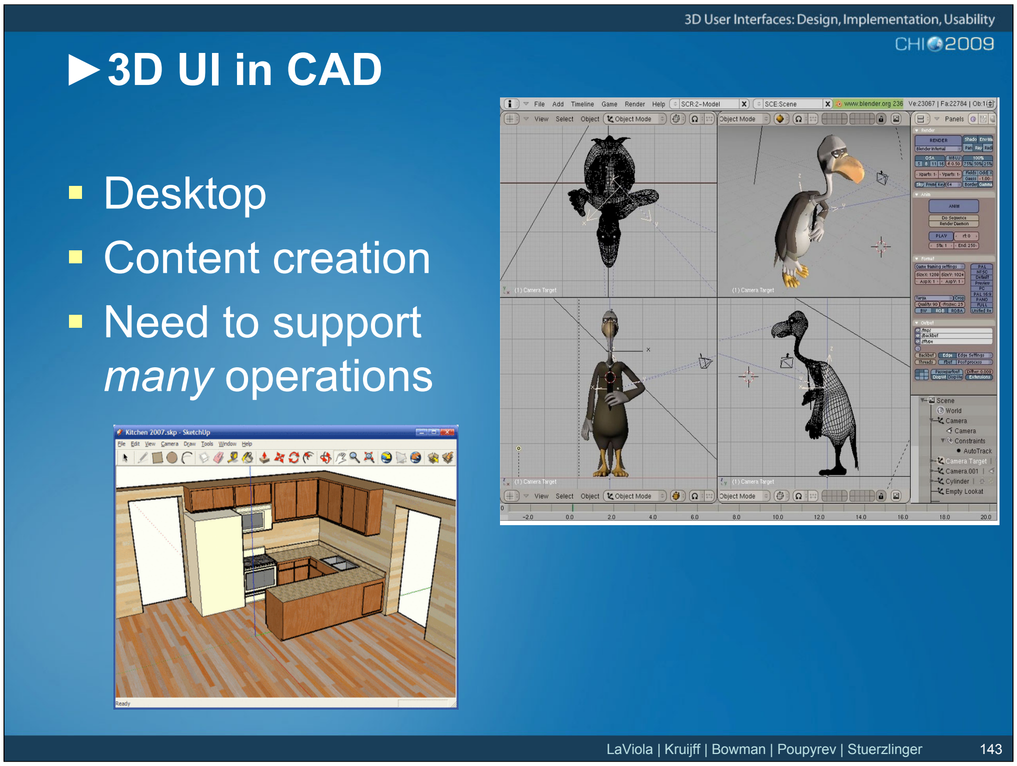
-Most people will never create content

Window/Desktop Managers

-Goal: Manage 2D & 3D applications

-Basically a means for “navigation” among applications, files, resources, etc.

-Desktop setting



Computer Aided Design (CAD), Animation, etc.

-Goal: content creation

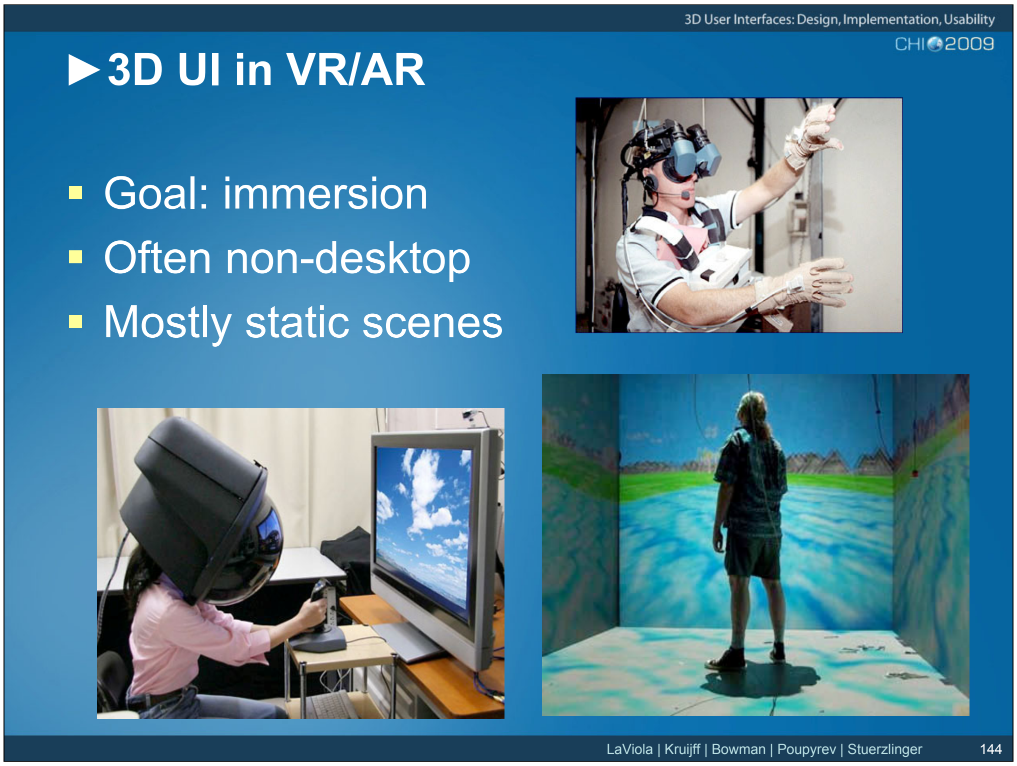
-Desktop setting, no stereo, a few 6DOF input devices

-Focus on manipulation

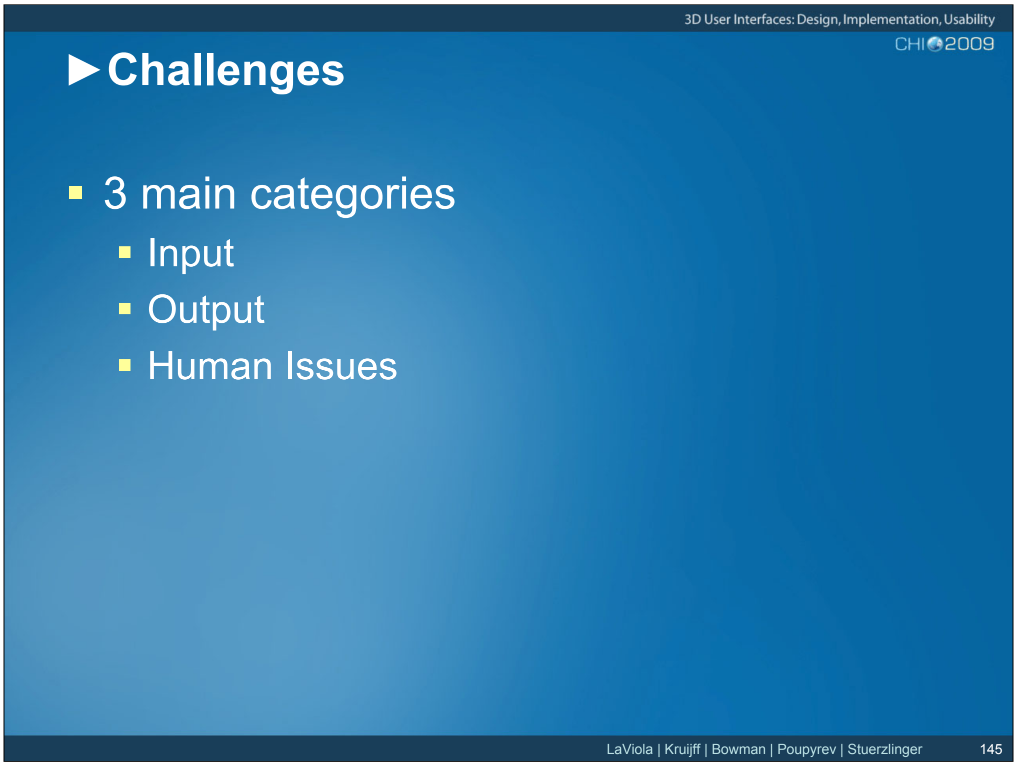
-Many, many geometric operations

-Hence very complex user interface

-Thousands of menus not uncommon



Virtual/Augmented Reality (VR/AR)
-Goal: increased immersion into 3D
-A plethora of input & output devices
-Walkthrough of predominantly static environment is the norm
    -Few interactive systems

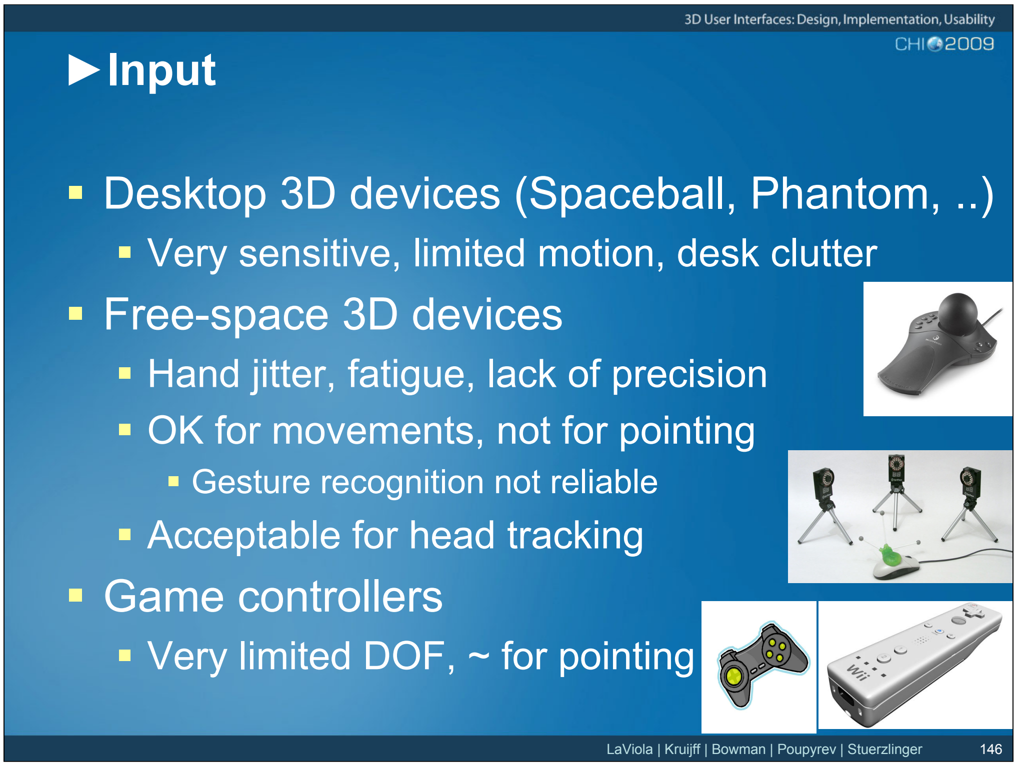


List of design considerations, loosely grouped into 3 categories: input devices, output devices, and human issues. Note that some of the issues clearly cross the boundaries of this classification.

Users can choose from wide range of input devices for 3D UI's. On one end are the classic desktop devices such as keyboards and mice. On the other there are advanced devices that allow users to control the position and orientation of objects in three-dimensional space directly. This is often referred to as 6DOF devices, as they can control six degrees of freedom simultaneously – three to control the position in space and three to control the rotation around each of the three axes of space. Finally, there are devices that afford control of either more or less than six DOF's.

Similarly, there are many output technologies for 3D UI's. They range from standard LCD monitors ubiquitous on desktops, to full 3D displays that create the illusion of content in a 3D space.

Last, but not least, are issues that are based on the capabilities of humans. Some parts are inherent to human nature, others are shaped by our environment. This part discusses those most closely related to 3D UI's.



In general, input devices are used in three kinds of settings. Either the device is used on a desktop, in free space, or in a game context (typically sitting on a couch).

On desktops, the combination of a (2D) mouse together with a keyboard is clearly one of the most prevalent interaction devices for 3D. Many CAD program use it as the sole means of interaction, where the third degree of freedom and rotations are accessible via modifier keys, mouse buttons, or various form of on-screen manipulators (handles). Most desktop games also use the mouse, typically by limiting the degrees of freedom of the user and/or the manipulation. However, there are also 3D input devices designed for the desktop. Among them are ball/puck shaped devices (Spaceball, etc.) that the user can grab and move in all three dimensions, as well as rotate in all three dimensions. As the ball/puck moves physically only about a millimeter or so, these devices are very sensitive to minuscule motions and this often leads to a negative first-use experience. Another class of input devices is small robot arms that enable the user to move a pen (or any attached to the end of the robot actuator) in a limited, approximately soccer-ball sized, volume of space. Practically all of these devices can also “push back” via motors in the joints, i.e. provide the user with a haptic experience (e.g. hitting the pen onto the surface of the object).

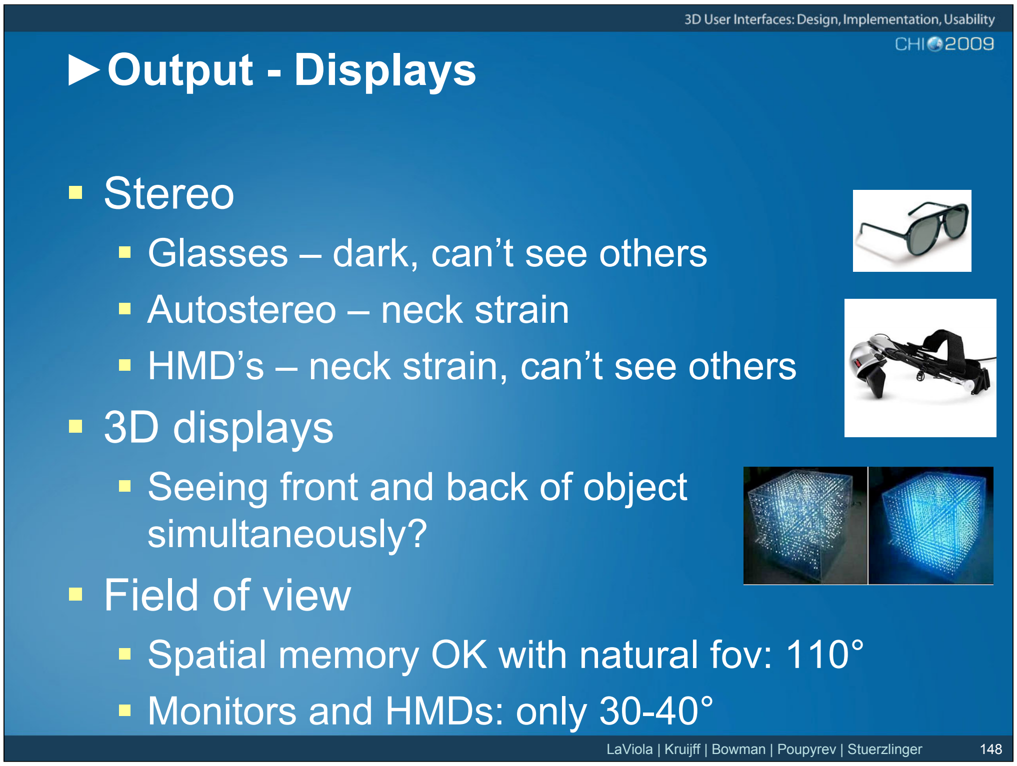
To track interaction in free space, usually multiple sensors that are placed overhead and/or around the user, which detect the location of the device via various technologies (optical, acoustical, electromagnetic, inertial, etc.). While most 3D tracking systems are good enough for Virtual Reality head-tracking (i.e. keeping the images aligned correctly for the current eye position of the user), they are usually not good enough for Augmented Reality work where the virtual images need to align with real world features. For interaction, free-space input devices suffer from technical jitter, hand jitter and user fatigue. Humans are simply not good at holding their hands *unsupported* in the air for minutes at a time. Moreover, the accuracy of almost all 3D tracking systems is at least one or two orders of magnitude less than that of a mouse, which has negative effects for fine manipulation.

In a casual game context, the user of frequently sitting on a couch. Here the predominant input device are game controllers, which typically afford separate control of independent degrees of freedom (usually 2x2DOF plus many buttons). One noteable exception is the Nintendo Wii remote (Wiimote), which can (somewhat inaccurately) track the position on the screen the device is pointing to as well as it’s distance. Moreover, it also adds 3D motion tracking via accelerometers, which can be used to recognize simple gestures with acceptable accuracy.

OK/Cancel comic on the idea of interacting in free space for extended periods of time.

Person coming out of a screening of “Minority Report”: “Mate, that film was brilliant! I recon that interface’ ll be the interface of the future!”

In the year 2099…: “I am sorry Ma’am. Your cognitive scores are incredible but you simply don’t have the upper body strength to do this 8 hrs a day”



There are many output devices that can be used for 3D systems. The ubiquitous desktop screens (LCD or CRT) are the most commonly used kind of display. They support only monoscopic display, but have high refresh rates and do not need calibration. Another cost-effective method to generate an image is to use a projector to generate an image on some screen or wall, although typically with lower resolution.

There are many technologies to generate stereo perception. The most common is stereo glasses, where different images are projected or displayed on the screen surface and the glasses separate the left and right image for the user. This typically necessitates twice the frame rate of normal displays. A big problem with all types of glasses is that they obscure the eyes of other users also wearing glasses (either due to polarization, filters or darkening). That means users can't see each other's eyes, which affects collaboration negatively, a fact well known from video conferencing. Also, most users prefer not to wear gear on their head for extended periods.

Auto-stereoscopic displays generate different images that can be seen from different viewpoints, e.g. for the two eyes of a human. This is typically achieved via some lenticular screen in front of the actual display. Most technologies require that the user hold their head stable in a relatively small region to achieve this effect. That leads to neck strain, which prohibits any form of long-time use. Head-mounted displays suffer from problems due to the added inertia on head movements, low resolution and a very small field-of-view for affordable models and are generally not used anymore (even the field of Augmented Reality is moving towards hand-held displays now). Last, but not least, there are true 3D display systems. There is a variety of technologies that can generate "glowing points" inside a volume. The main issue with this concept is that users then see the front *and* back of objects simultaneously, something the human visual system is not trained to interpret easily. One new class of systems that have been demonstrated recently in the lab generates different images for different viewing directions, similarly to the auto-stereoscopic systems. These technologies allow the viewer to move freely, but are not yet at a stage where they can be used in office or home settings.

One particularly insidious drawback of smaller field-of-views is the lack of peripheral vision. This affects navigation and/or spatial memory negatively. Single LCD/CRT monitors, head-mounted displays, and stereo glasses with thick frames all suffer from this problem.

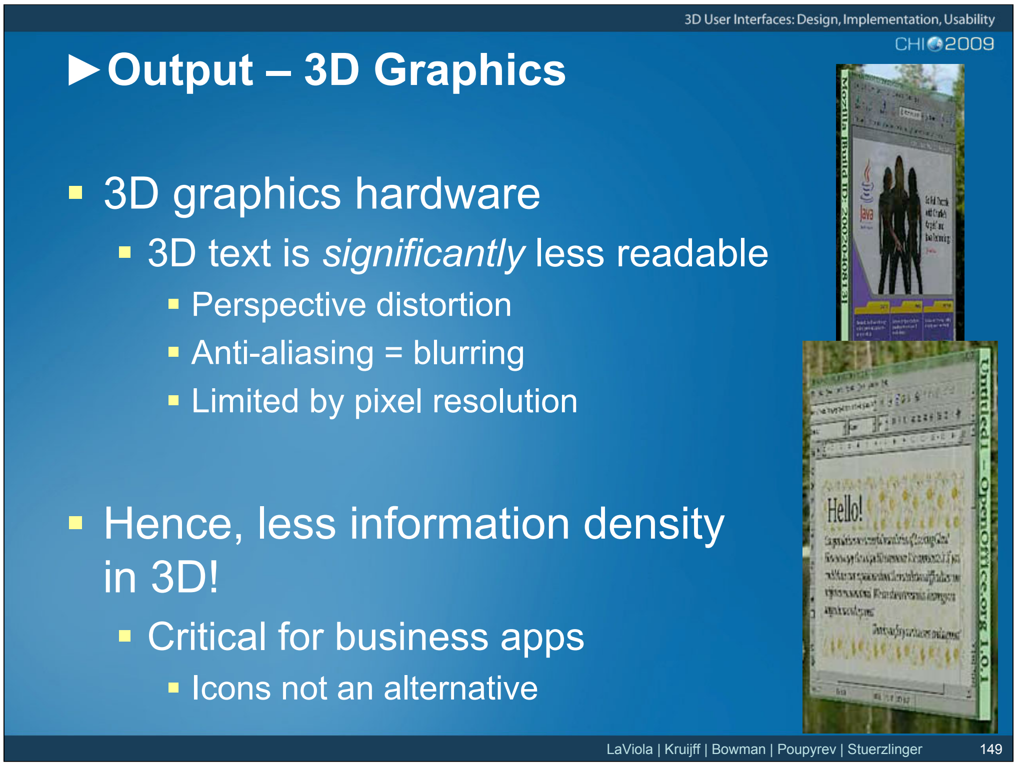


Image generation for 3D interactive systems normally involves use of 3D graphics hardware. Great advances in performance and image quality have been achieved here. Very important for 3D interactive systems is that text is significantly less readable when displayed in 3D compare to 2D. There are two interdependent factors that cause this. One is that *single* words become less readable when they are rotated around a vertical, even with optimal, yet prohibitively expensive, anti-aliasing. “Normal” anti-aliasing methods increase the blur but this decreases readability The other issue is that the perspective distortion causes large parts of the text on a page to become extremely small – often smaller than a pixel. Imagine a page of text rotated around the vertical axis by 45 degrees so that the left side of the page is closer to the viewer. In this example the beginning (left side) of each line is easily readable, but the text at the end (right side) of each line is too small to be readable. All other alternatives (e.g. top, bottom, right side of page closer to viewer) are equally bad or even worse.

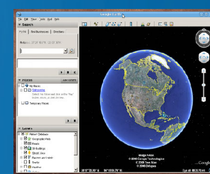
Hence, we can say that 3D text can only be displayed with significantly less information density compared to 2D text. In other words, the amount of information that can be shown on any given display device is significantly less if 3D text is used. However, information density is critical for many real-world applications. One often-proposed alternative to increase information density in 3D environments is to use “recognizable” 3D icons. However, automatically making “good” icons for content is a very hard problem that has currently no real solution. Another potential solution is to substantially increase the pixel resolution of current displays, e.g. to 600dpi, but this has prohibitively high bandwidth requirements.

First, and foremost, humans are not "naturally" proficient at full 3D navigation, i.e. (more or less unconstrained) 6DOF movements. Evidence for this is that people in "full" 3D professions, such as astronauts, divers, fighter pilots, usually need extensive training to do their job. Astronauts also need training because they work in an environment without gravity, and they actually have to "un-learn" their reliance on gravity. Moreover, most human environments are not fully 3D, nor do they require full 6DOF navigation. Mostly people constrain themselves to 4DOF (walking in the plane and looking around). Tilting the head is unusual, and changing the height of the viewpoint is usually accommodated with a complete change of posture. As another indication, consider the contortions of a plumber to be able to see a connection under a sink – many people prefer not to do this.

As for manipulation, humans are good at manipulating an object in 6DOF if they can grab it with the fingers of *both* hands simultaneously. However, if they can use only one hand, this ability is much reduced as e.g. rotations of 360 degrees become impossible without "clutching". Even worse, if a person has to manipulate an object indirectly via a handle (as is typical in 3D user interfaces), performance drops even further. If, and only if, some form of support or contact surface is available, humans leverage it to simplify manipulation. Witness e.g. the prevalent use of scaffolding in many professions concerned with creating or modifying small or large structures. Another example is a plumber who uses the rigid connections betweens pipes to construct three-dimensional structures. In summary, humans are not necessarily as proficient in full 6DOF tasks as one may believe.

▶Human Issues 2

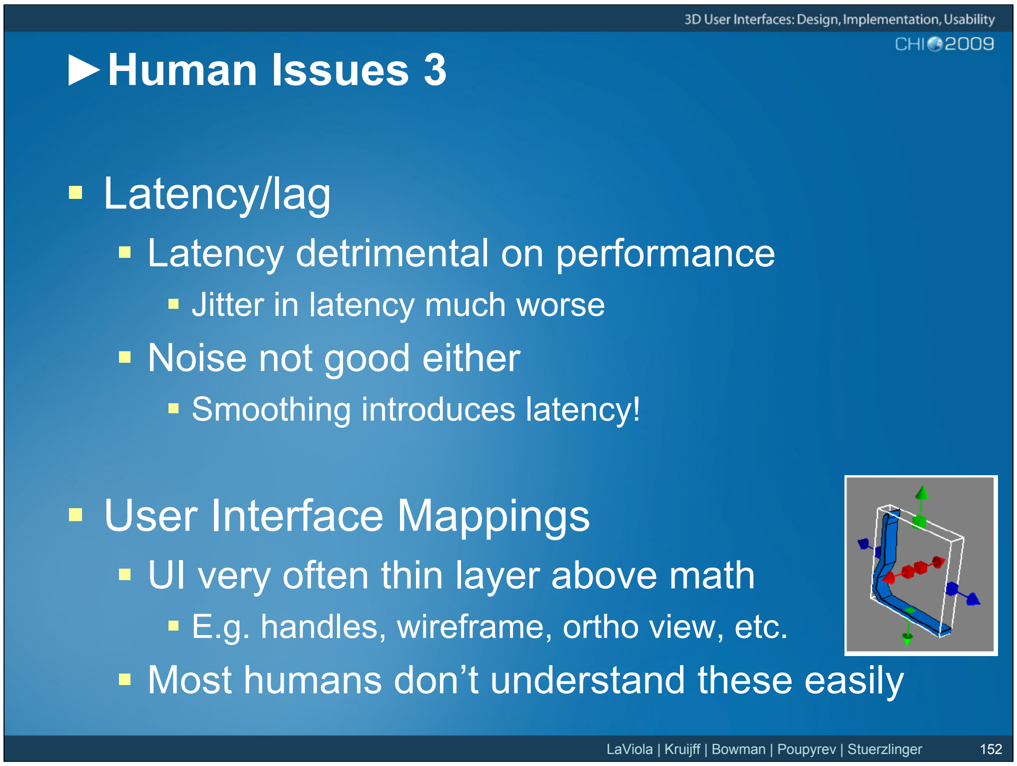
- People interact only with *visible* objects
  - Strong preference
- Depth perception not that accurate



- Navigation
  - 3D spatial memory not much better than 2D
  - Easier/faster to teleport/search
    - Google Earth

Humans also prefer strongly to interact with objects that they can see directly. If part of an object is invisible, people will either rotate the object or move themselves to see it before working on it. Again, the way plumbers' work is an excellent example here. In other words, manipulation of invisible objects is the exception, not the rule. Related to this is the fact that depth perception of humans is relatively less accurate compared to the accuracy across the visual field.

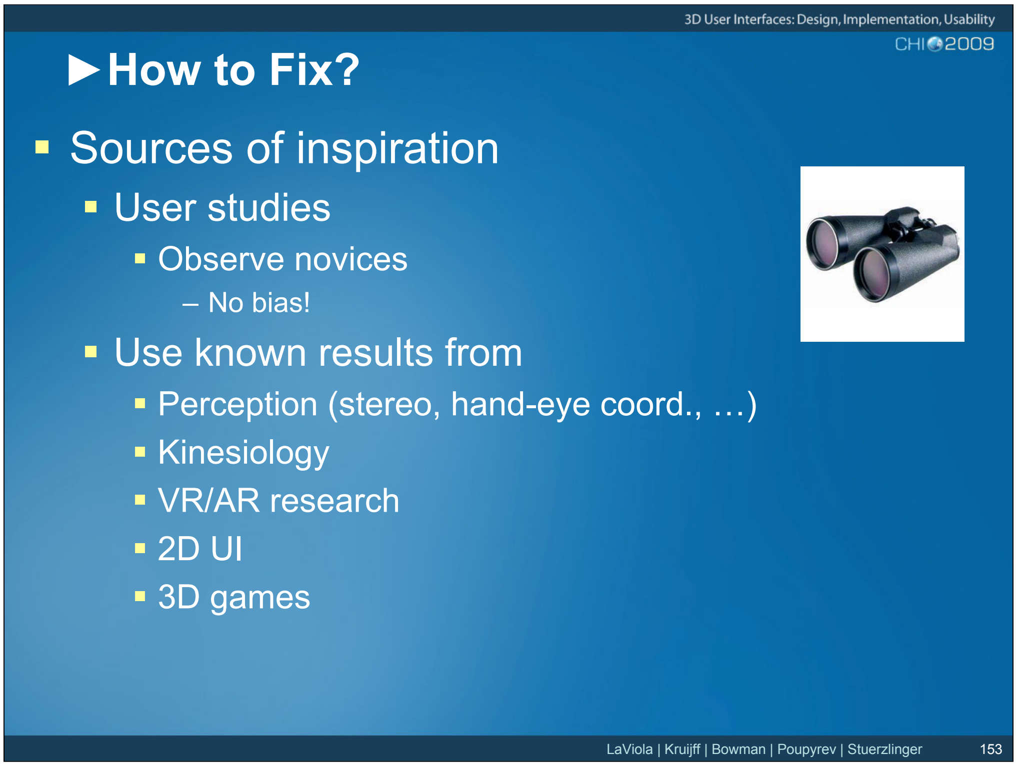
Moreover, 3D spatial memory is not *that* much better than 2D spatial memory. The main reason for this is that the world is only a restricted 3D environment. Consider e.g. that buildings have *numbered* floors, connected by elevators and stairs. Hence, most humans remember the floor number and the 2D location on that floor, but not the spatial location in 3D. Similarly, furniture has drawers or doors that are only accessible from the front, which forms again a 1D or 2D indexing system. And objects are organized inside the drawers to simplify access, too – very frequently in a 1D or 2D layout. Finally, consider that although systems such as Google Earth afford 3D navigation, most people navigate only within a very small region. Larger travel is usually handled by "jumping" to a new location, either via search or bookmarks. This is an indication that people prefer to "teleport" for larger distances rather than navigate.



Yes another property of the human “system” that affects both input and output devices simultaneously, is that people are sensitive to latency or lag. In other words, any delay in the handling of movements (regardless if it is on the tracking or on the display side) has negative effects on human performance. Measurements have shown that even delays as small as 16 milliseconds may be noticeable and do affect performance. While can adapt to constant latency to some degree, they still rate it negatively. However, any *variation* in latency has usually disastrous effects.

Another common issue is the effect of noise in tracking which affects performance, too. Only as long as the noise is significantly smaller than the smallest necessary movement, there is no real problem. Moreover, if targets are very small (e.g. because they are far away in a perspective view), any amount of noise will affect performance. The technical alternative of smoothing out the noise seems like a good idea, until one realizes that smoothing introduces extra latency – which in turn *decreases* performance too!

Last, but not least, we have to consider how “natural” user interface mechanisms are. Consider e.g. that engineers need training to understand wireframe views or orthogonal projections. In other words, such mechanisms are not appropriate for the average person. Or consider that 3D handles that move objects along the coordinate system axes or planes require that the user has an understanding of the concept of local and global coordinate systems – again something that is not natural for most people. Finally, many computer-aided design systems offer manipulation methods that are a one-to-one mapping of the underlying mathematics or a very thin layer above it. In this case the user needs to understand the mathematics to be able to use such a system effectively, which is often not practical. Consider e.g. how difficult it is to put a crease into a NURBS surface in current computer-aided design systems.



Given all the challenges mentioned above, we need to look how we can go forward

As inspiration, we looked at many different sources:

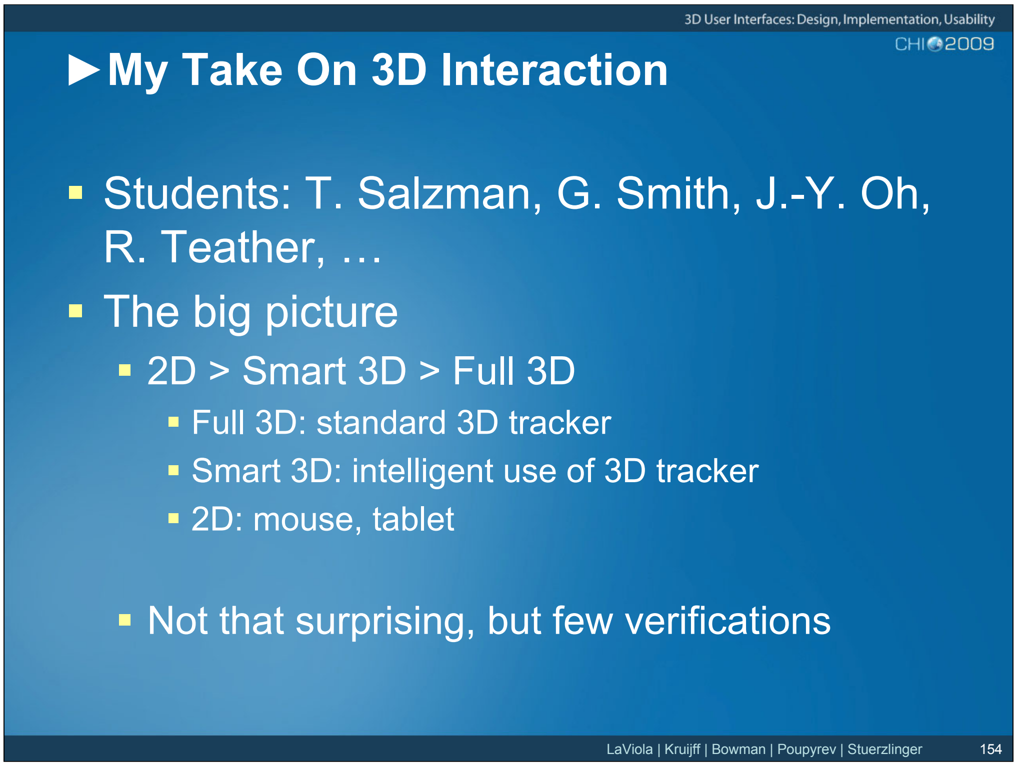
-In our lab, we performed many different user studies on 3D UI's over the years. Typically we use participants without 3D knowledge (3D CAD, VR/AR, … experience) in our studies, as few people have this kind of training. Of course we cannot completely remove the known effects of 3D games, as a large part of the population has been exposed to this.

-Research into human perception, both for the visual system as well as the hand-eye system (typically in kinesiology)

-The literature on user interfaces in VR/AR research.

-The rich body of knowledge on 2D user interfaces. While not all lessons learned there are directly applicable, many guidelines for this field still apply also to 3D user interfaces. E.g. consider the guidelines for good visual design, guidelines for error prevention and recover, etc. For brevity of this presentation the reader is referred to the Human-Computer Interaction (HCI)/user interface literature.

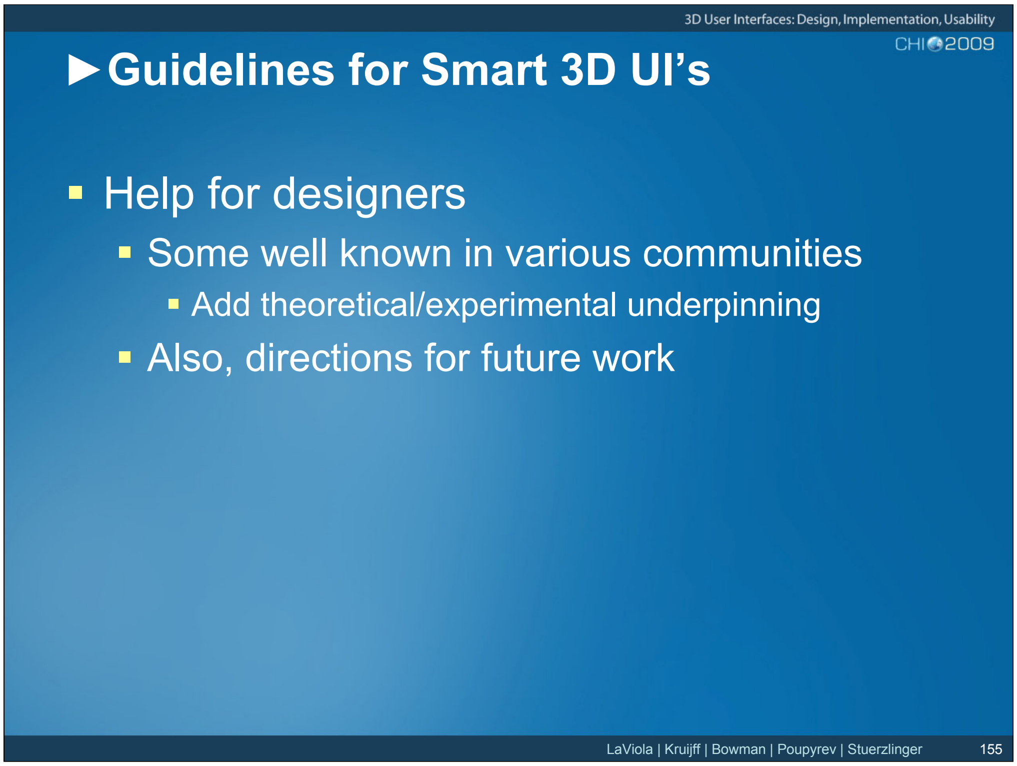
-3D games, as they also explore the space of 3D UI's and only those techniques that are "good" survive for any length of time in this rapidly evolving market.



Much of the following is based on the work done in my lab, with the most important students listed.

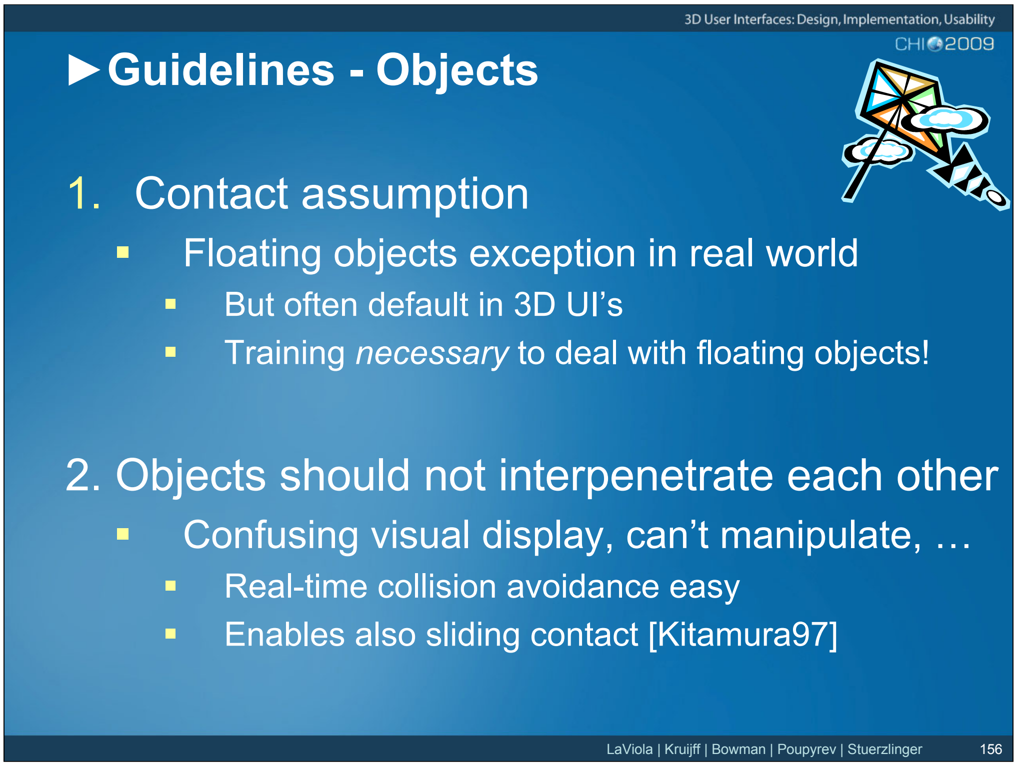
Overall, 2D interaction is better/faster than “intelligent” 3D user interfaces, which in turn is better/more efficient than user interfaces that rely of full 3D interaction

While few people disagree with this statement, few verifications have been done. More to the point, few people have analyzed why 3D user interfaces are slower. This is one of the areas that my lab works on.



To help designers of 3D user interfaces, we will now present a list of guidelines for good 3D user interfaces.

Not all of these are new, and are well known in various communities. To help the field, we have collected the most important ones in one place. Moreover, we have added theoretical and/or experimental underpinning to the guidelines.

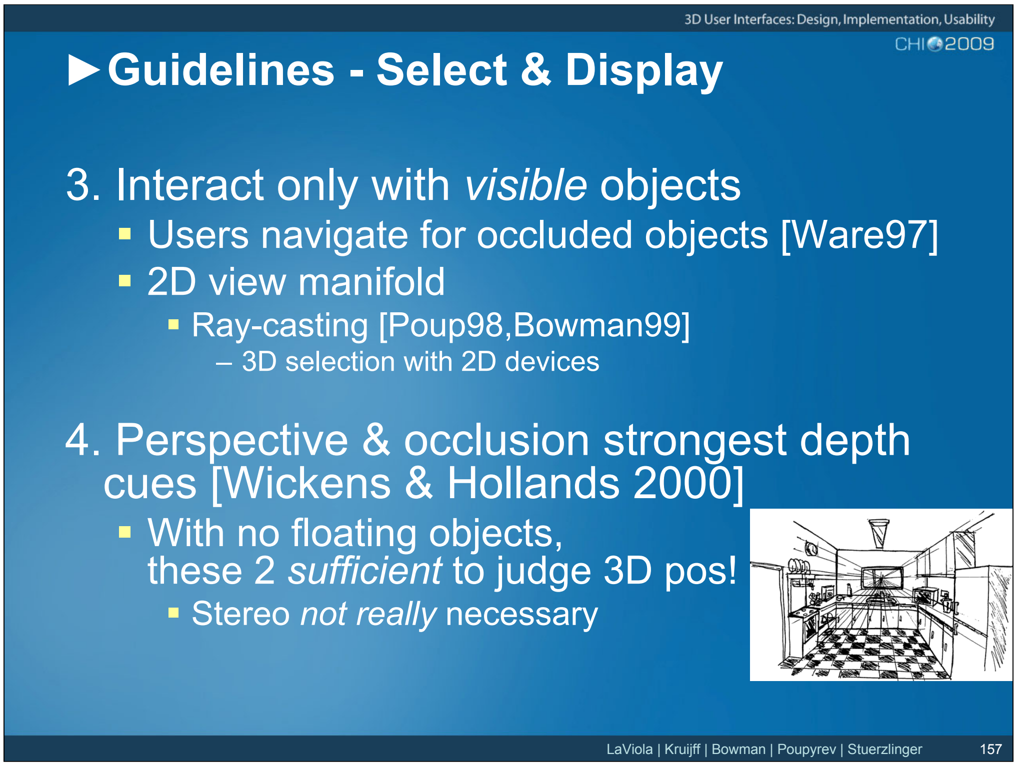


Contact Assumption

In the real world, few floating objects exist, and almost all objects are attached to other objects. However, the default in most 3D systems is that every object floats. To leverage humans' experience in the real world in a better way, the right *default* for a 3D system is for objects to attach to other objects, if only due to gravity. Interfaces can and should incorporate special mechanisms to make objects stay in midair.

Objects Don't Interpenetrate

Solid objects — including the viewer him or herself — can't interpenetrate each other. Humans are used to this and deal with it every day. However, many VR systems allow object interpenetration by default. Interpenetration leads to confusing visual display, and many novice users can't easily recover from such situations. For example, consider the negative effect of users being "trapped" behind a wall in a game — most novices need help to recover from such a situation. Today, performing real-time collision detection and avoidance for large environments is fast with the help of graphics hardware. As an added benefit, collision detection and avoidance enables sliding contact, one of the most efficient ways to position objects in the real world.

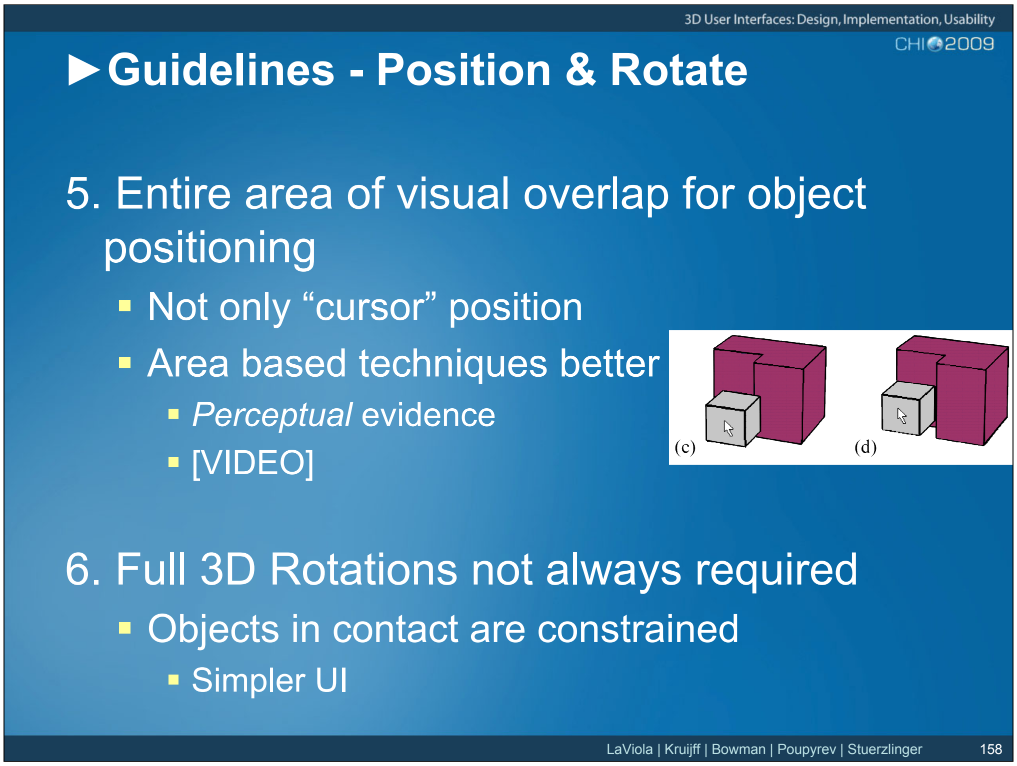


Interact Only with Visible Objects

Users interact with what they see. As such, humans will navigate so as to see or better see objects before interacting with them. This is even more more important when the 3D environment has no tactile/haptic feedback. This has several consequences. First, it points to the importance of easy navigation. Second, because a 2D manifold can fully describe the set of all visible objects, 2D input is then sufficient to select an object! This is also documented by the success of ray-casting and occlusion based techniques relative to point-based virtual hand techniques. This also means that 2D input devices are at least sufficient to select objects in a 3D world — assuming that adequate 3D navigation techniques exist.

Perspective and Occlusion Are the Strongest Depth Cues

For manipulation of objects beyond arm's length and assuming a static scene and viewer, perspective and occlusion are the strongest depth cues. Assuming that there are no floating objects, these two cues are usually sufficient to accurately and quickly judge objects' 3D position in an environment, unless optical illusions are involved. Although stereo display has a clear value, it matters most for objects fairly close to the viewer. Consequently, stereo display of 3D environments isn't always necessary. Last, but not least, evidence exists that most stereo technologies are far from mature and are tiresome or problematic if used daily, up to and including cyber sickness symptoms.

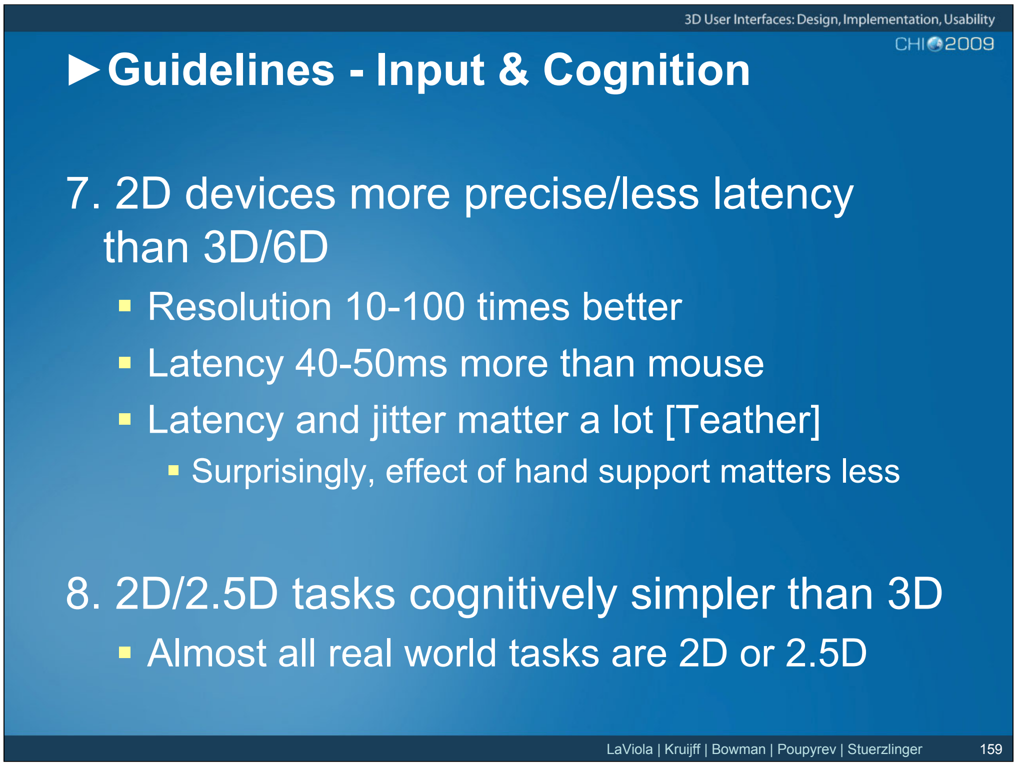


People See the Object, Not the Cursor

Research into primate vision has demonstrated that monkeys attend visually to not only the tip of a tool in their hand but also the whole tool and the hand. This indicates that a cursor might not be the best choice for 3D Uis — a cursor is effectively a point, while an object covers an area in the visual field. The Sesame (Sketch, Extrude, Sculpt, and Manipulate Easily) sliding technique analyzes the visual-area overlap between the manipulated object and the static scene to determine a moving object's position. The user studies reported in conjunction with this research demonstrate that users can easily use and learn such techniques and that such methods provide clear performance benefits.

Full 3D Rotations Aren't Always Necessary

Many common objects, such as chairs, desks, and shelves, have a clear "up" orientation. Other objects, such as hanging lamps and whiteboards, also have clear orientations. These objects are all attached to other objects. This attachment also provides constraints for rotation — a chair is on its side only in exceptional cases. Consequently, providing a simple user interface to rotate an object around the axis afforded by that object's main attachment is a good design alternative for simple-to-use systems. Although the interface should support full 3D rotations, such modes can be secondary and don't need to be easily accessible.



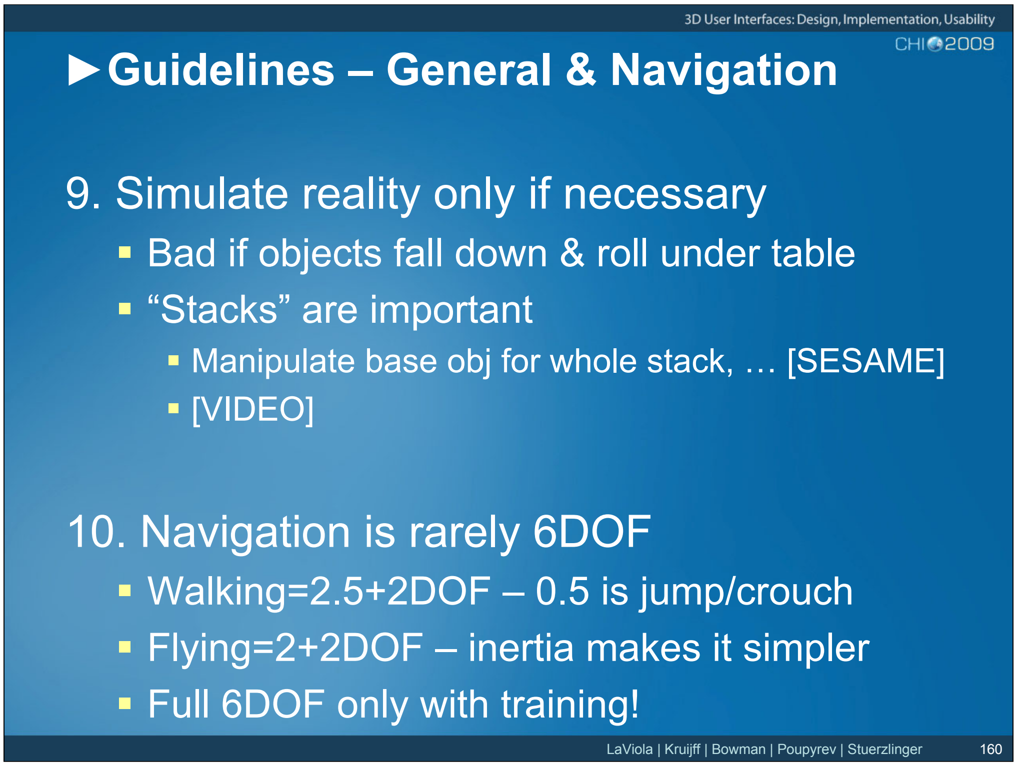
2D Input Devices Are Advantageous

Input devices such as the Personal Interaction Panel, which use a pen on a 2D tablet to provide interaction in a VR system, have been shown effective for 3D worlds. Also, constraining the input to 2D combats hand fatigue and provides more accuracy. Moreover, a comparison of input device specifications between mouse- or pen-based systems and 3D technologies reveals that 2D technologies are one to two orders of magnitude more precise. This comparison also showed initial evidence that this precision difference is one reason why 2D input devices outperform 3D technologies, with another difference being the increased latency of 3D tracking technologies compared to a mouse. Interestingly, a supporting surface's effect is much less than that of increased resolution. Consequently, combinations such as using a tablet PC with accurate pen tracking along with a 3D tracking system for off-slate interaction are a sensible approach.

2D Tasks Are Cognitively Simpler Than 3D

Most real-world tasks aren't fully 3D; they are 2D or 2-1/2D as the extra information is not useful. For example, blueprints of buildings abstract the height dimension so as to better focus on 2D spatial relationships and multistory buildings are layers of 2D floor plans. When needed, cross-cuts show alternate dimensions or perspective drawings show 3D. Real 3D structures in buildings exist, but they are again the exception, not the rule. Consequently, most humans are used to dealing with 2D or 2-1/2D and don't have the training necessary to deal with problems that are fully 3D. Yet another example here is the way stacks of objects (paper, clothes, cards, etc.) are handled. People quickly learn that one can't just pull an object out of a stack. Instead one has to lift the top of the stack away to reveal the desired object then work with that object and finally reassemble the stack. One example for this is that the SESAME system analyzes the scene structure which affords quick and easy manipulation of such stacks.

In summary, this means that offering 2D methods to achieve most tasks is an excellent way to increase usability for 3D user interfaces.

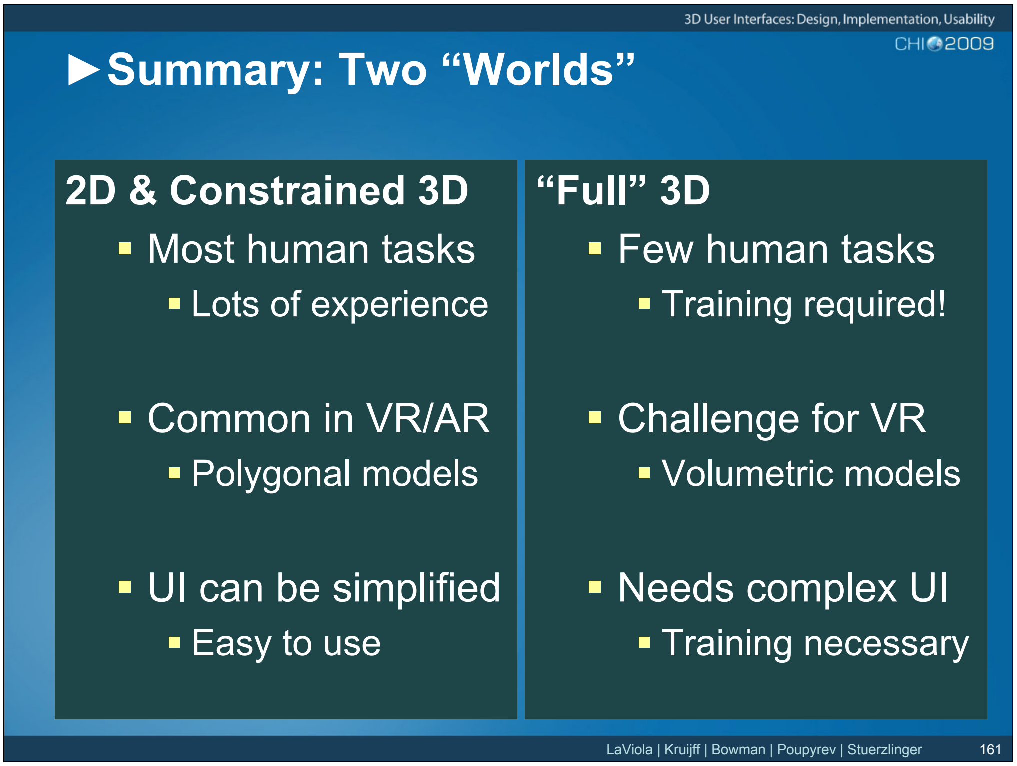


Reality Simulation Isn’t Always Appropriate

One potential option for 3D user interfaces is to simulate reality more or less completely. However, besides being technically challenging, this is not appropriate for many applications. Consider e.g. an object being bumped off a table and rolling under a cupboard. Retrieving that object is cumbersome and not necessary in a 3D user interface – unless the application focus is on the retrieval task.

Constrained Navigation And Teleportation Is Good

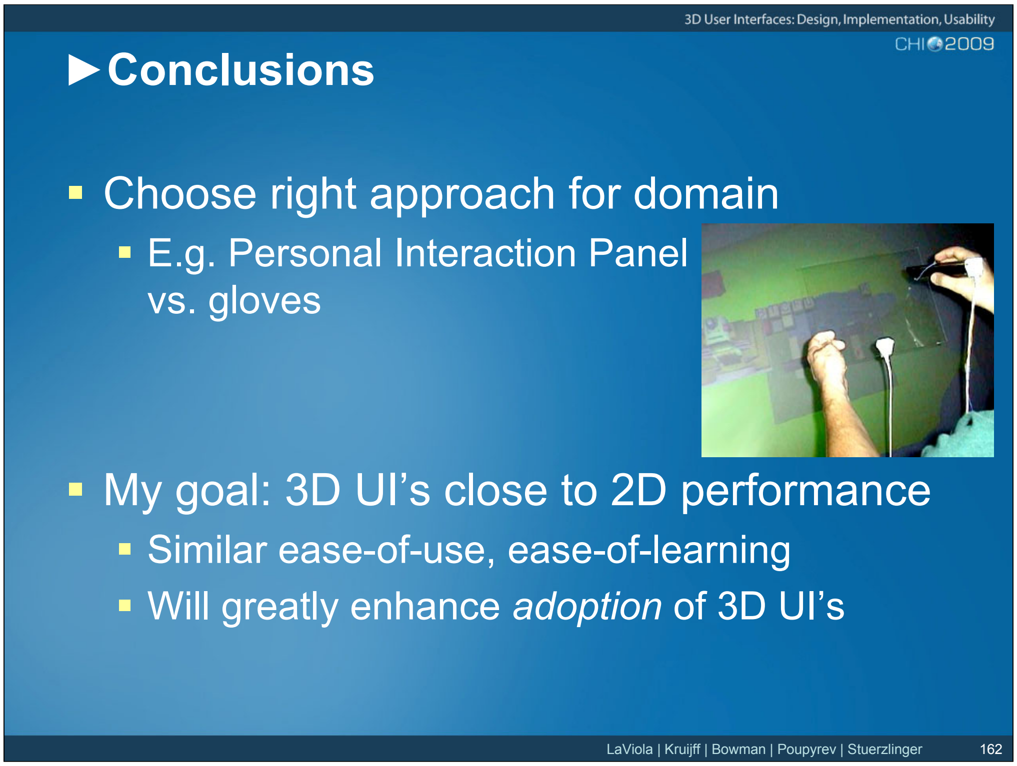
In the real world navigation is rarely unconstrained, i.e. requires manipulation of all 6 DOFs. Even a fighter pilot is limited in their navigation capabilities because an airplane has effectively only 4 DOF’s and is moreover subject to inertia. Helicopter pilots require even more training because they have to control even more degrees of freedom. In general, most navigational tasks have 4 or less DOF’s, which can be used to simplify the user interface. Moreover, as navigation for larger distances is cumbersome, many systems provide a means of instant transportation to different location. This is usually associated with a search feature that allows users to specify a name for a location.



A different way to look at what we have discussed is to point out that there are 2 different worlds:

First, there is the 2D/constrained 3D world. This corresponds directly to many human tasks and experiences. Most VR/AR, game, and CAD environments as well as window managers fall into this category. Surface-based representations are abundant in this area, and user interfaces can use these surfaces to simplify operations. The most important characteristic of this area is that full 3D user interfaces are not needed here and/or have negative performance characteristics.

Second, there is the "full" 3D world. While this is a very desirable target, humans are generally not well-equipped through their experience to deal with the challenges. Hence, typically a lot of training is required. Moreover, this area involves tasks that require operations in free space and/or inside volumes (i.e. are not surface-based). Here, more complex user interfaces are required, which in turn require training. However, for all operations that are constrained, simplified user interfaces are still beneficial here!



One way to summarize this lecture is to state that designers of 3D UI's should use the right approach for the domain in question. E.g. the "Personal Interaction Panel" is a better choice for many common tasks in high-end VR systems compared to interaction with gloves.

Overall, the goal of my work is to create 3D UI's that are easy-to-use and easy-to learn. However, we first need a solid understanding of the issues that make current 3D UI's difficult.

Wolfgang Stuerzlinger

Department of Computer Science & Engineering / Centre for Vision Research

York University

Toronto

Canada

http://www.cse.yorku.ca/~wolfgang

The following people are acknowledged for contributing:

Chad Wingrave, Univ. of Central Florida

Rob Teather, York

Ji-Young Oh, Autodesk

Graham Smith

Tim Salzman