What is

COCODEX?

V 0.18

Jaron
Lanier

International Computer Science Institute,
Berkeley, CA

Lead Scientist, National Tele-immersion Initiative

please note: if you are reading the pdf version of this document, the movies can be found at
http://www.advanced.org/jaron/pentagon/cocopent15.ppt_media/

Preliminary research supported by: sgi

Part One:
Introduction

These are the goals of the Cocodex project:

Create a personal communications and computing interface device with many of the advantages of a dedicated advanced command/control center, that can also support visual telecommunications **better** than any existing command/control center, provide key VR/3D capabilities, and that is portable, cost-effective, rapidly deployable, and has a small footprint.

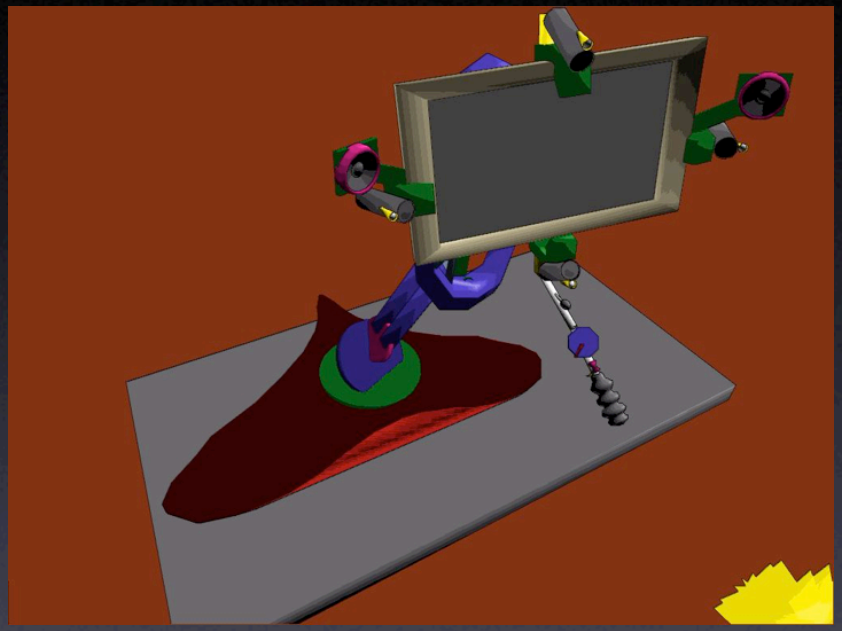
Although COCODEX addresses problems in the domains of command/control, visual tele-communications, and rapid deployment, the easiest introduction is via comparison to well known Virtual Reality devises...

The two primary instrumentation strategies for Virtual Reality are the Head Mounted Display and the CAVE

Mid 1980s VPL HMDs

Fraunhofer Institute of Industrial Engineering

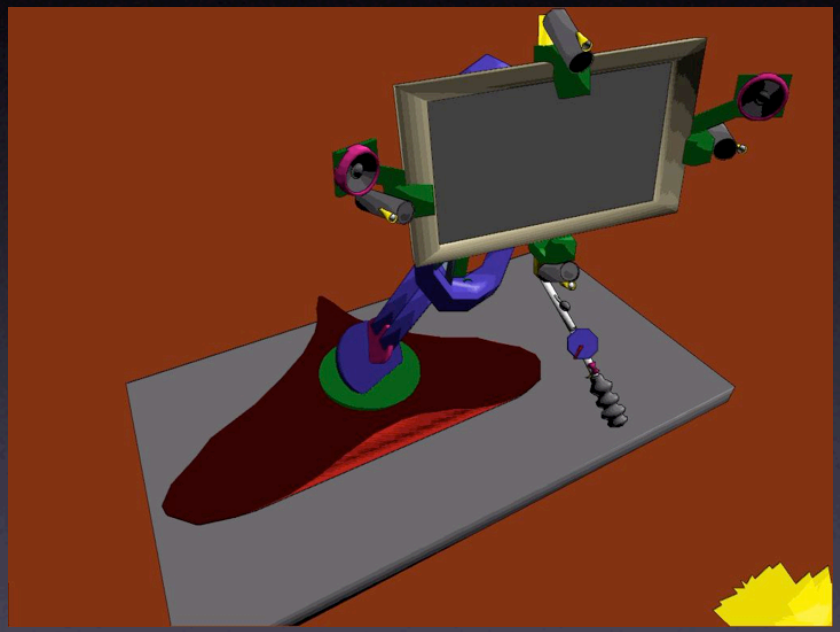
COCODEX is the halfway interpolation between these two designs.

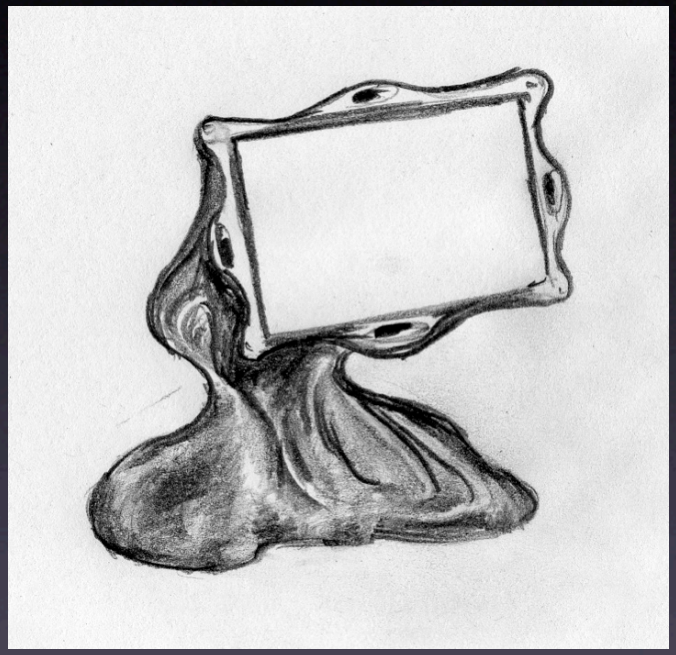


The way a sunflower follows the sun...

COCODEX is an array of sensor and display elements mounted on a robotic arm that follows your head around without touching it.

The illustrations of COCODEX in this presentation depict near term designs that can be built to test the ideas.





An eventual commercial version would probably be lightweight, soft, and biomimetic, for reasons to be explained.

COCODEX can be thought of

- as a way to simulate access to a command center, CAVE, or display wall on a desktop,

- as a thus-far unique solution for making full duplex tele-immersion possible,

- as a thus-far unique strategy for lowering requirements of transducer quality so that existing cameras, displays, and other parts are already good enough,

- as a way to reduce bandwidth and latency requirements for tele-immersion, and

- as a way to use immersive and non-immersive user interfaces at the same time.

It will take some explaining to introduce all of these applications of the COCODEX design!

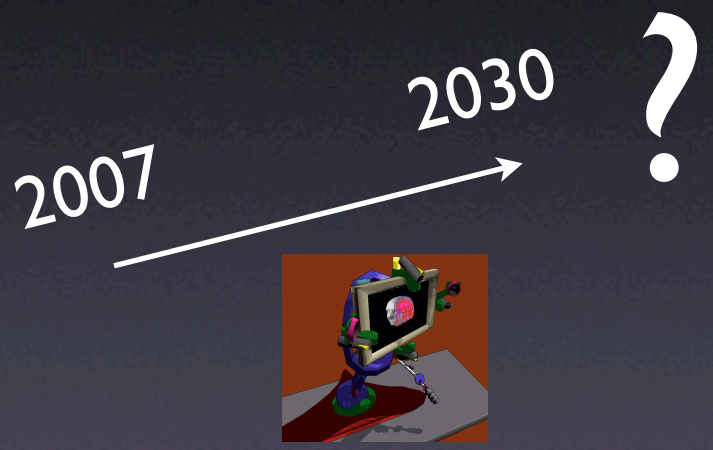
The key to understanding COCODEX is in examining details of its control structure. Please take the time to consider these details. COCODEX is a design with subtle qualities that has the potential to solve a range of important long-standing problems.

Specifically, there are two major long-standing problems addressed by COCODEX:

- Full duplex tele-immersion
- Display real estate crunch and the "Special Room" Dilemma

If COCODEX solves these problems it will vastly expand the usefulness of information technology.

Eventually, ambient universal sensing and in-eyeball or in-brain displays, or who knows what, will probably be a better solution. Cocodex is a model of a solution using already known components that could last a few decades...



A summary of COCODEX strengths and weaknesses is found at the end of this presentation (slide/page 108.)

Compact
Collaborative
Desktop
Explorer

(Also a type of medieval songbook compilation.)

Current Academic Collaborators

- OS Software

  Oliver Staadt, UC Davis

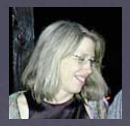
- Mechanical Engineering/Haptics

  Kenneth Salisbury, Stanford

- Human Factors/CogSci

  Jeremy Bailenson, Stanford

And... The fakespace guys, Mark Bolas and Ian McDowell; Lenny Lipton of Stereographics, Mary Lou Jepsen, Hartmut Neven, and other assorted characters



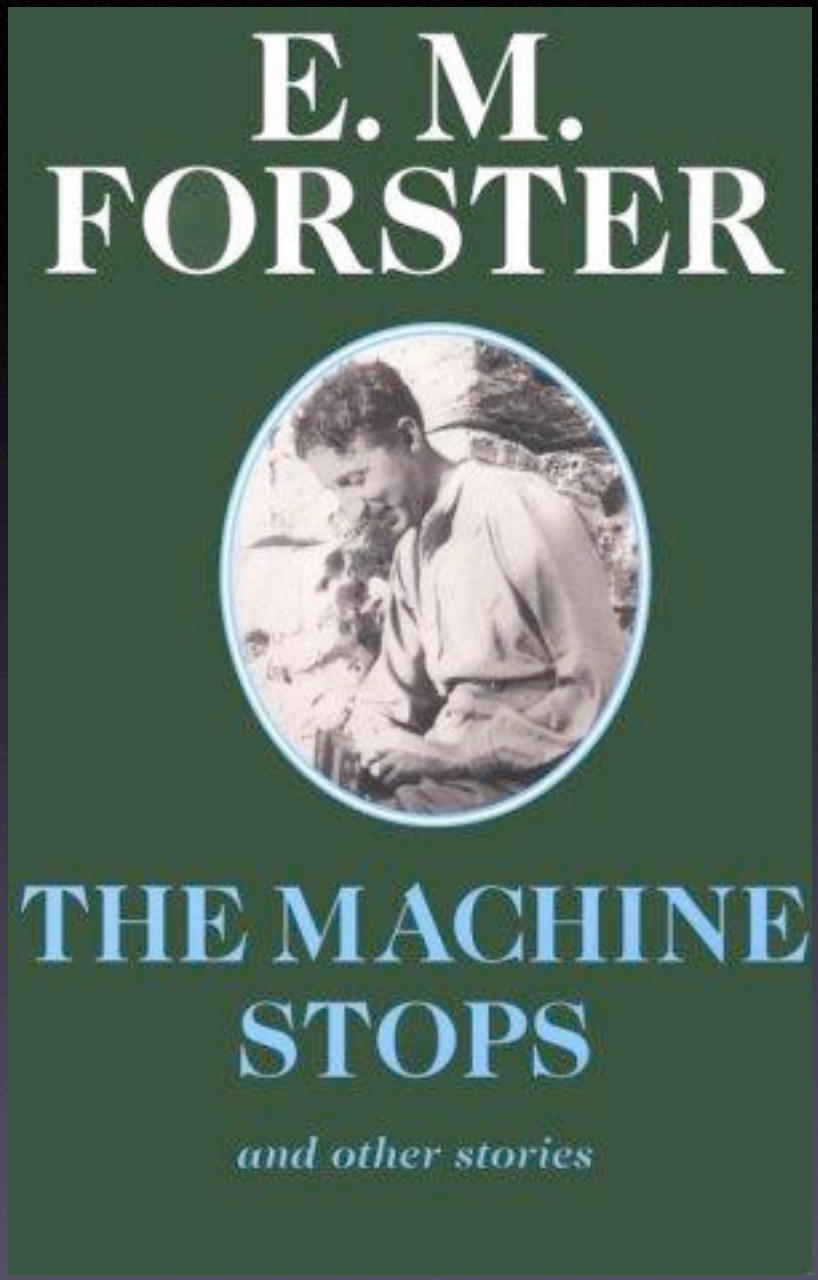
Part Two:
Why the tele-immersion problem is important, and why it hasn't been solved yet.

(This section contains background materials only. If you're familiar with Tele-immersion and advanced UI research, you'll want to skip to slide/page 39.)

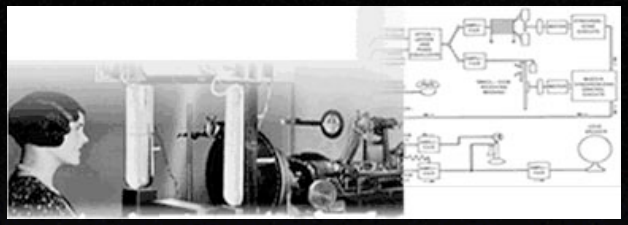
"If, as it is said to be not unlikely in the near future, the principle of sight is applied to the telephone as well as that of sound, earth will be in truth a paradise, and distance will lose its enchantment by being abolished altogether."
Arthur Strand, 1898

1909...

"But it was fully fifteen seconds before the round plate that she held in her hands began to glow. A faint blue light shot across it, darkening to purple, and presently she could see the image of her son, who lived on the other side of the earth, and he could see her."



Video conferencing has always looked great on paper...



This image quality was described as "perfect" by the New York Times reporter who covered the first demo.

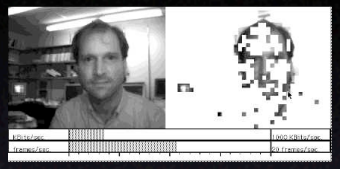
1927

NEW LOOK PICTUREPHONE

Now you can *see* as well as talk

The Picturephone has Touch-Tone controls to make calls and control the television screen so you can see the person you're talking to, be seen yourself, or have a darkened screen. Attended service between New York, Washington and Chicago began in 1964.

1964



CUSEEME: Tim Dorsey, first image

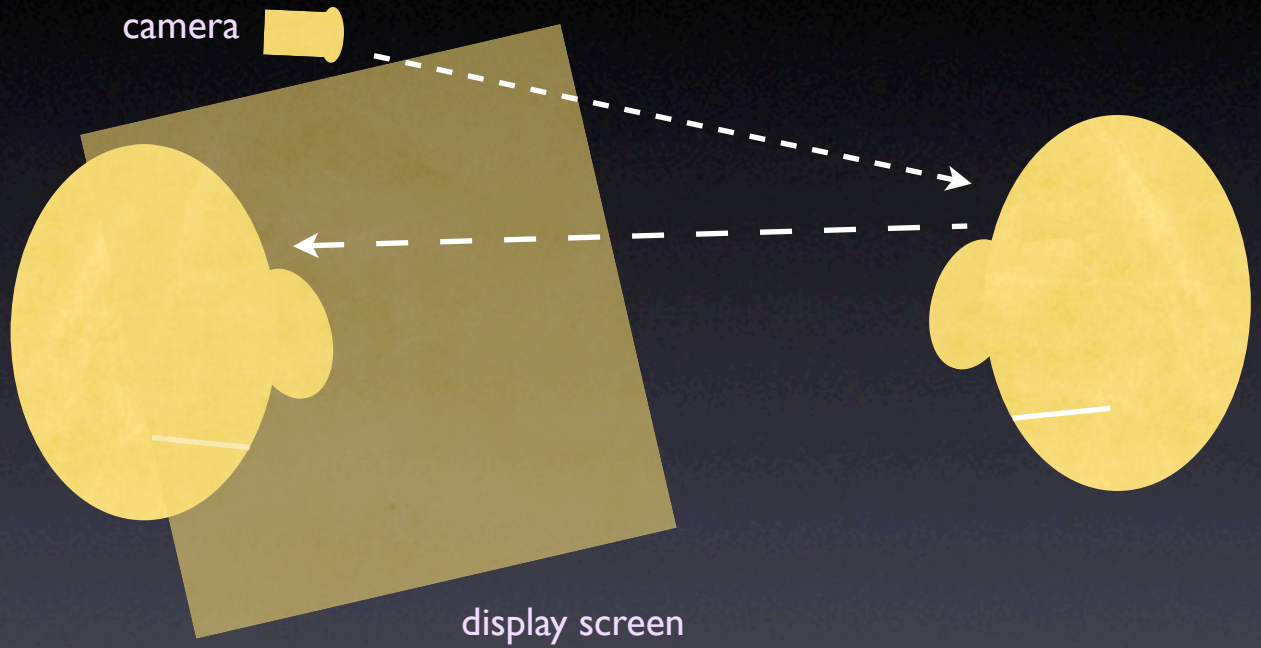
INRIA

1992

NetMeeting

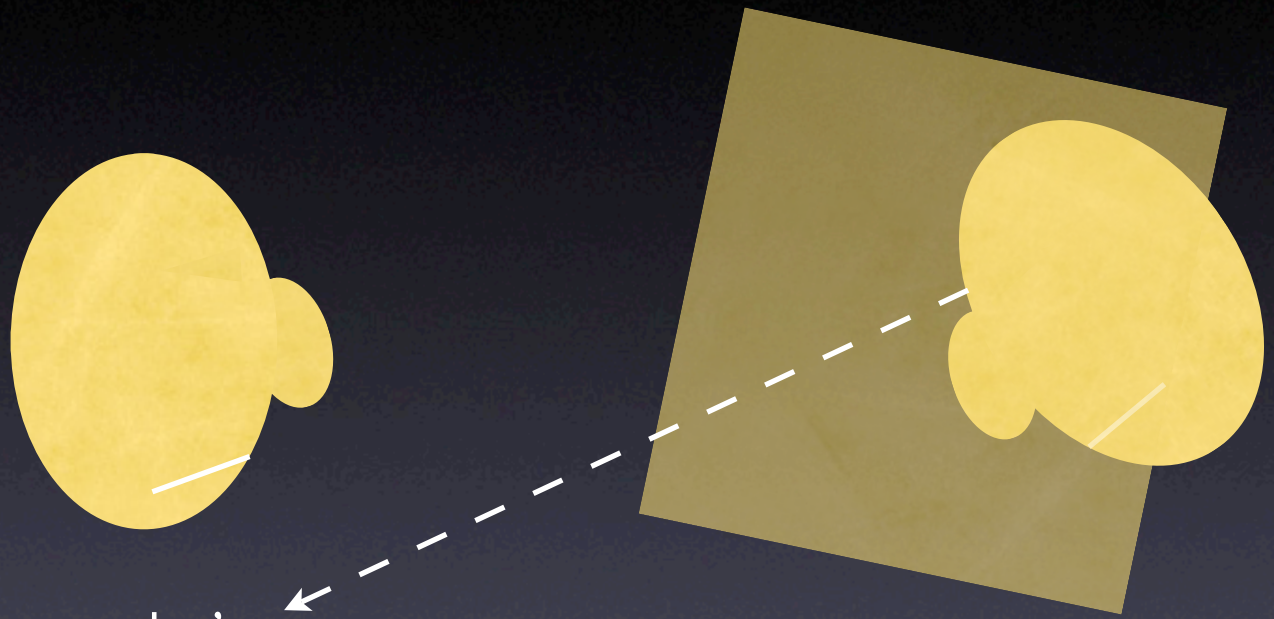
current

But human factors issues have never been resolved...



The celebrated eye-contact problem

try as they might, users appear to each other to be looking away



you don't want to know...

One class of solutions applies to two participants sharing a single sight line on a virtual axis

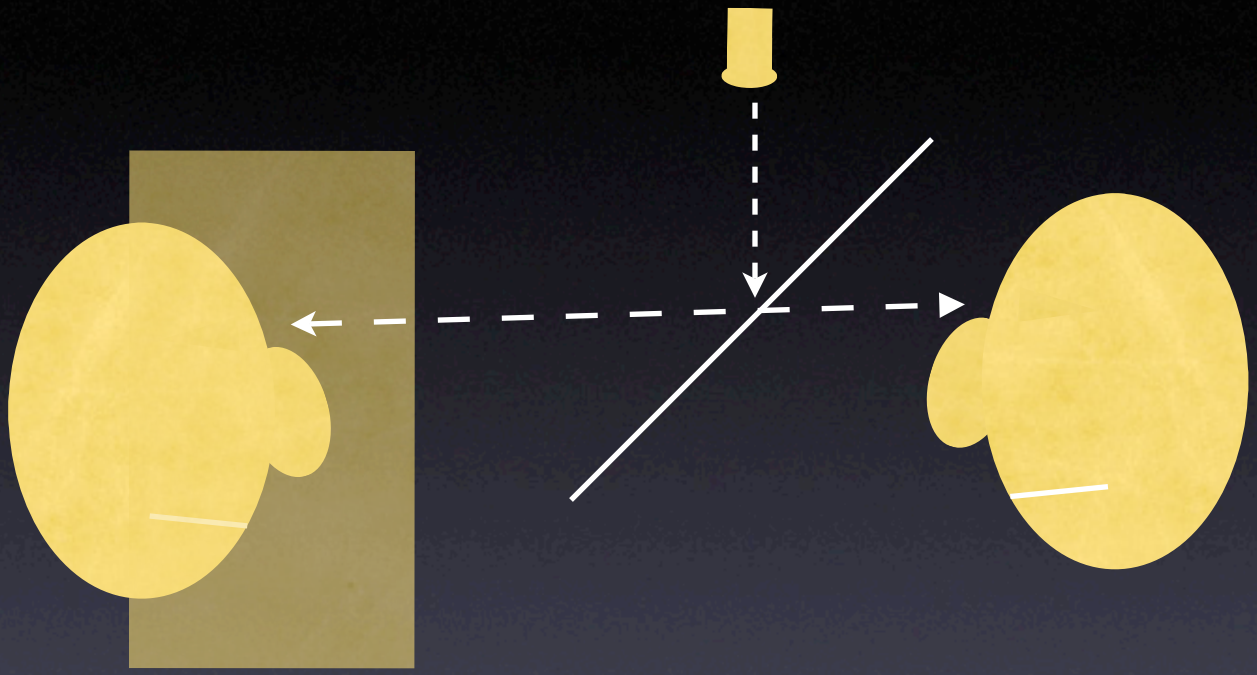


image-based version at UNC

Eye-catcher product from Exovision (NL)

you can achieve this with a half-silvered mirror, or image-based simulation of same, or screen with camera elements in tiny holes, or many other variations- dozens of patents a year in this solution class for last two decades

This solution class breaks down in real world use, however.

- >2 Participants a common requirement.
- The normal range of motion in conversation precludes a constrained sight line. Restricted motion degrades usability.
- A single sight line can't support two eyes as accurately as needed (although there are still honest disagreements on this point.)

Most common current strategy is to skew images of people so that a correct perspective isn't even suggested, and to restrict resolution so that cues, especially related to eyes and mouth, are ambiguous.

docomo phone

telesuite

ichat mockup

"HYDRA", Bill Buxton's 1980's approach to the >2 users problem

polycom

typical campus roll-your-own using assorted standards

barco

All the above configurations share this strategy.

Fundamental civilian demand drivers for better telecommunications

- Peak Oil

  (Don't expect nuclear commercial aircraft anytime soon.)
- Globalization and Outsourcing

  (Collaborators everywhere.)
- Distributed Families

  (Long distance elder care.)
- Post-Napster Economics

  (Personal interactive contact more valuable than bits.)
- Hopefully terrorism will lose its place on this list.

Military applications summary:

- Rapid deployment and redeployment of advanced command/control stations without requiring construction or decommissioning of dedicated facilities.

- Survivable distributed advanced command/control in the event of an infectious agent WMD attack (which would preclude the gathering of personnel into command control centers in violation of quarantines.)

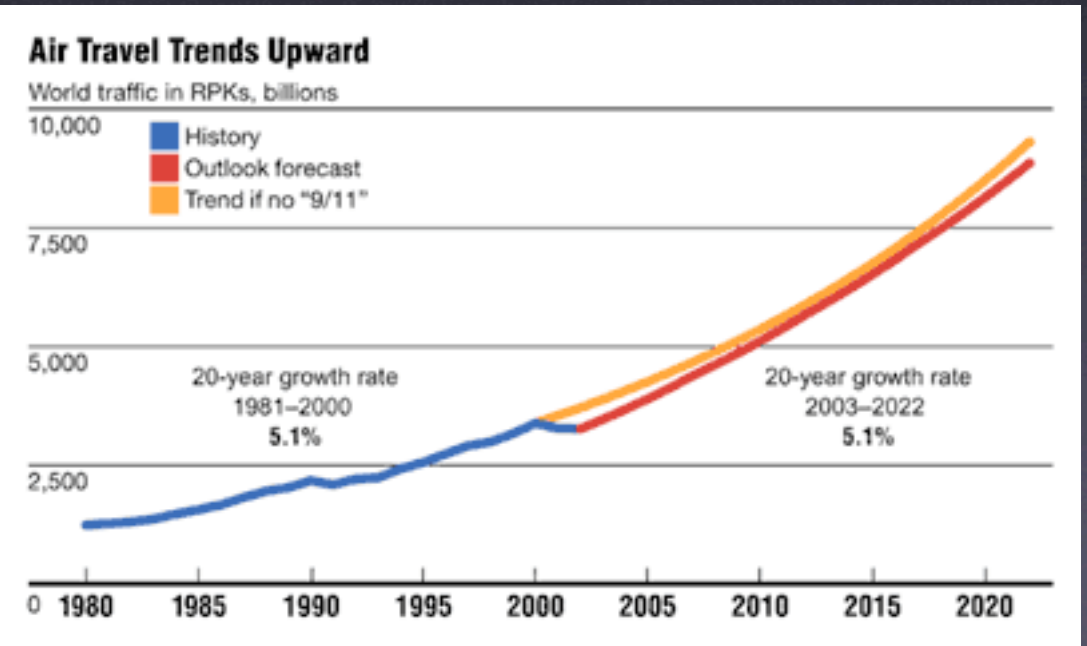
- Potential improved command communications due to improved visual tele-communications.

- Remote presentation of advanced 3D information without dedicated facilities.

So...

Why do people buy plane tickets so often instead of relying on the telephone, email, or video-conferencing?

Source: Boeing attempting to sell stock



Something's missing from tele-communications technology as we know it.

Humans have been optimized by evolution to perceive other humans well (since other humans were a primary threat to survival and the only source of mates, and childhood learning was profoundly expanded.) Thus realistic digital presentation of humans to each other is a profound challenge.

Media technology succeeds or fails relative to expectations set up by intrinsic patterns of use.

TV and video-conferencing started at the same time, but only TV took off.

Why does low res television work better than video-conferencing?

Because:

a) Interactivity sets up higher expectations for video-conferencing.

b) Television and movies benefit from celebrity: We integrate internal models of the people we see over time, and the lack of interactivity protects our illusions.

Eye-contact is only the best known of a long list of troubles with video-conferencing. Another prime example is latency.

Why does the telephone work better than video-conferencing?

Because:

a) Pauses for breathing can be perceived as being of ambiguous length, thus masking latency. If we had a separate orifice for speaking, intercontinental phone calls might not work!

b) An absent sensory modality is less troubling to users than the conflicting cues from crummy multi-modal implementations.

How severe are the ultimate human factors requirements?

No one knows. We must use a spiral strategy in which new generations of instrumentation allow ever more refined tests. The last few generations of the spiral seem to be converging, so we are probably getting close.

"Virtual Reality" explored a new approach to visual telecommunications.

Ivan's rig at U Utah, 1969

VR was originally defined as multi-user extension of Ivan Sutherland's "Virtual World" including virtual bodies of the users so that they can see each other.

VPL networked immersive avatars, 1989

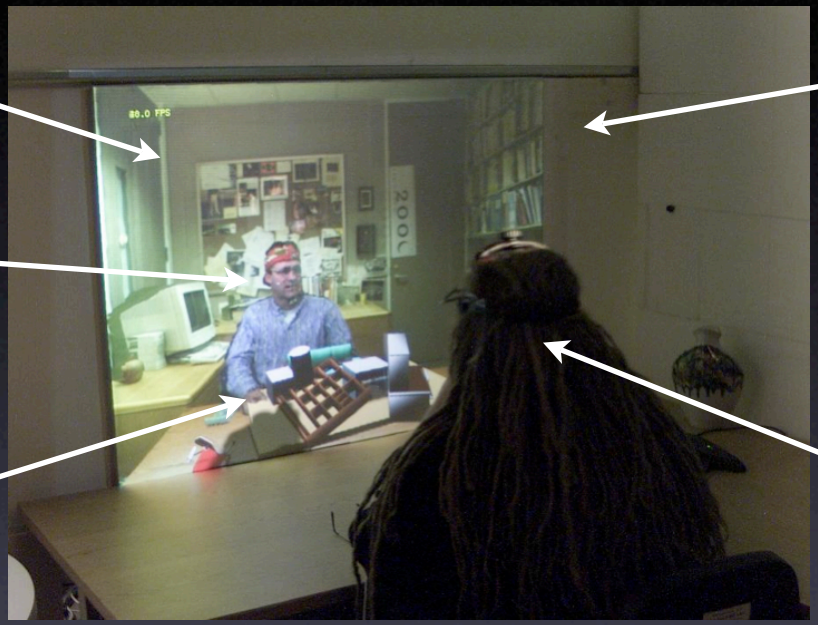
First full body avatar, Ann Lasko, 1987

"Tele-immersion" is loosely defined as the application of VR techniques to teleconferencing, or as tele-conferencing that solves human factors problems.

Remote location is Brown U in Rhode Island

Autostereo display for proper sight lines (will explain shortly!)

This is a dynamic real time volumetric reconstruction of Robert Zeleznik, my remote collaborator.



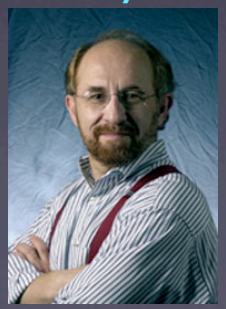
These are virtual CAD objects we collaborated on.

I was physically in Chapel Hill, North Carolina

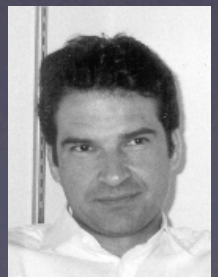
Oct, 2000

The first major Tele-i research project was the National Tele-immersion Initiative of Internet2 in the 1990s

Henry Fuchs



Kostas Daniilidis



Ruzena Bajcsy

Andy van Dam

Al Weis, Patron Saint

Key Co-PI's

Overview of history:

video-conferencing → human factors roadblock

virtual reality → facilities/ instrumentation roadblock

tele-immersion → happy medium?

(This is a movie)

1920s | 1980s | 1990s | finally?

Tele-immersion can support >2 users with correct sight lines.

New York

Rhode Island

2000

North Carolina

Another major Tele-i research project was “blue-c”

CAVE made of giant shutters
so camera array can see in

volumetric data from camera
array

end result

blue-c unites multiple CAVE users

Yet another Tele-i project was HP's Coliseum

2002

2003

An experiment in keeping the visual sensor array large enough to allow viewing from multiple collaborators, while keeping the screen small (and giving up true sight lines.)

Unfortunately, none of the implemented ideas in Tele-immersion instrumentation can simultaneously support full duplex communication, >2 users, and the normal range human motion while seated.

This person (a volumetric version of Amela Sadagic, as it happens) can't see properly, but posed this way to make the picture look good...

Many of the of configurations fail to achieve full duplex because users must see each other with stuff on their faces!

Quick survey of the full duplex problem:

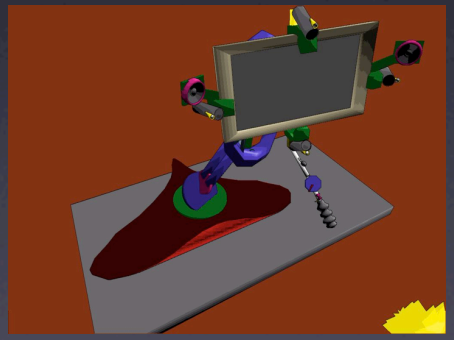
a) If you use a CAVE or HMD, there's stuff over your eyes, so it's hard for the system to sense how you should look to another person. CAVE glasses can almost be made to look like regular glasses, but not quite to the point of likely widespread use.

b) You can almost synthesize in software what a person would look like without headgear, but not quite.

c) If you don't have stereo at all, you could eye-track a person to provide an averaged, or "third eye" line of sight, but that approximation probably isn't good enough for sustained use. (There are still honest disagreements on this point.)

d) If you don't have any stuff on your face but each eye still sees an accurate scene for its position, that's called autostereo. Autostereo only works well enough for tele-immersion if your head doesn't move much relative to the screen, so as we'll see, cocodex makes it practical.

What will the next generation of tele-immersion instrumentation look like? How close are we to solving the problem?



No one knows the answer for sure, but the COCODEX proposal provides one potential answer.

There have been some devices that bear physical resemblances to COCODEX

It would be a little like an older imac...

Also a little like the "Boom Chameleon"

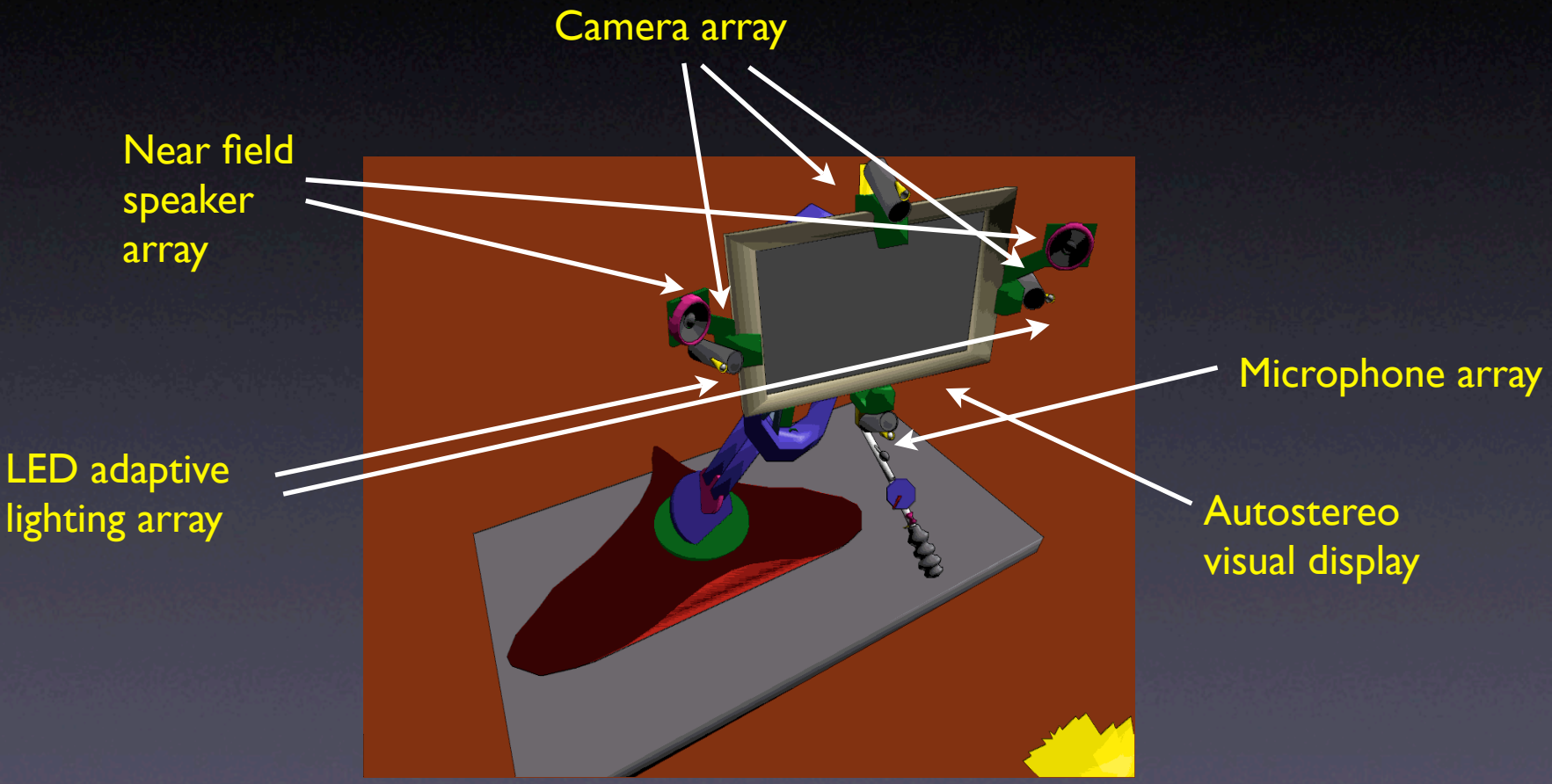
(Tsang, Fitzmaurice, Kurtenbach1,Khan, Buxton; all Alias/SGI, U Toronto)

Also a little like the various "hair dryer" schemes for VR from the 1980s that never quite worked...

But what distinguishes COCODEX from all these devises is a unique control structure.

Part Three:
What's so special about the control structure of Cocodex?

In order to understand the unique control structure of COCODEX, let's start by looking at some of the components of COCODEX's mobile I/O array.

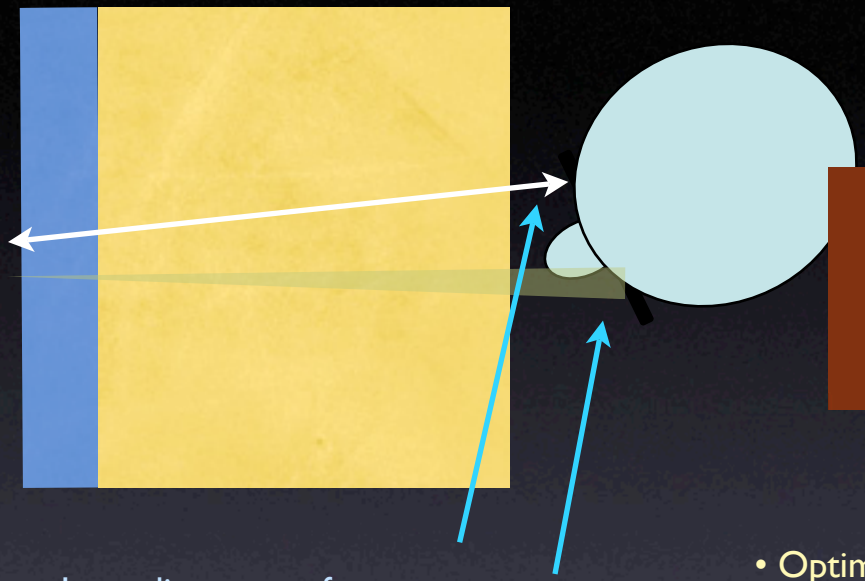


Every one of these subsystems exists in some form already. As it happens both the existing and near-term predictable versions of all these subsystems work well enough only so long as a person stays within untenable spatial and angular constraints.

Cocodex provides a unique escape from this dilemma.

CAVE
(or other big wall screen)

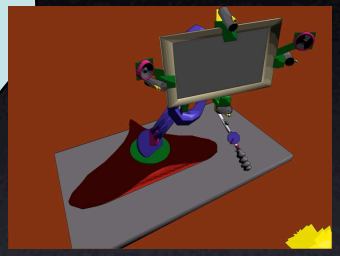
Too far?



- Long distance to face.
- Hard to get enough resolution with available cameras.
- Hard to place enough cameras in array to get coverage of range of motion and sufficient angles of observation at the same time.
- Small errors in angular alignment of cameras are amplified.
- Hard to get enough display resolution.

COCODEX

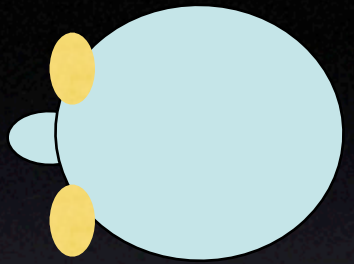
Just right?



- Optimal distance of available cameras to face.
- Useful camera array placement.
- Reasonable display properties with available components.
- No contact with face, so facial pose is neither obscured nor distorted by use.

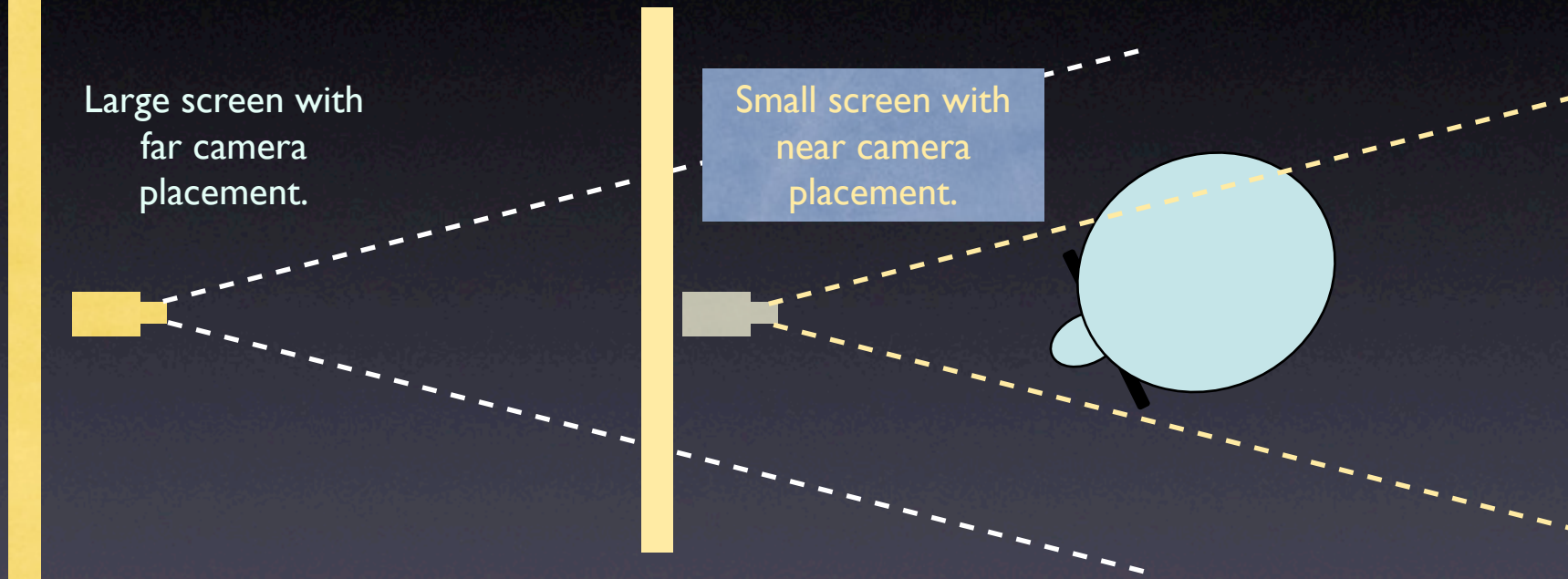
HMD

Too close?



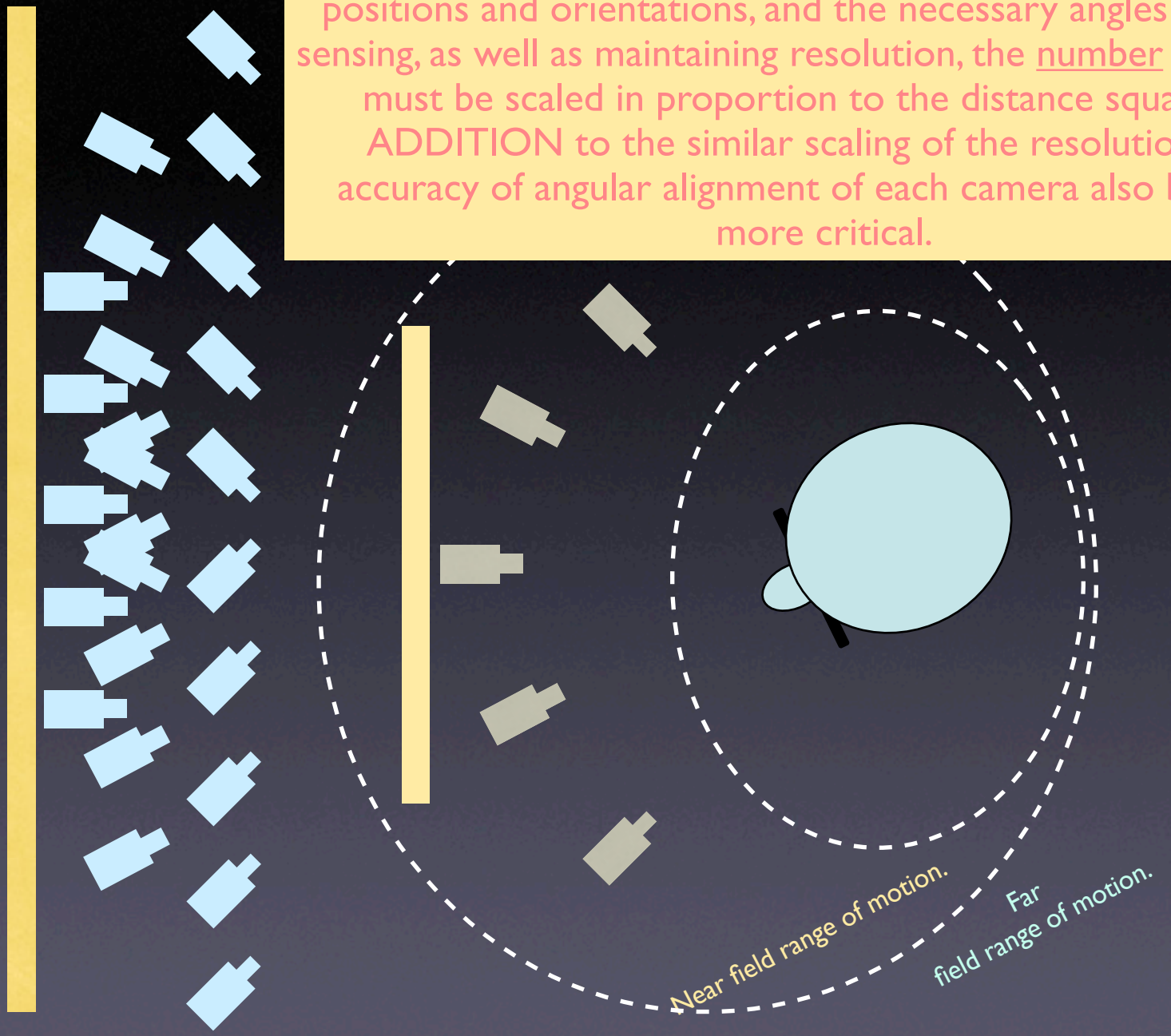
- Device not only covers, but distorts face.
- Hard to get good enough resolution and angle of view in display.
- Hard to get enough camera distance to measure certain facial events within HMD.
- Small shifts in how HMD sits on head amplify facial expression tracking errors.
- Variations in head size, hair style, and other factors make ergonomics difficult.

Note that even though large displays are getting better and cheaper, the sensing side of a large display will lag behind by perhaps a decade. Here's why...



A single camera must see an increase in actual resolution proportional to the square of distance to maintain effectively constant resolution if only one point of view is needed, but as we've seen, effective visual tele-communications makes an even greater demand.

In order to cover both the expanded range of possible head positions and orientations, and the necessary angles for visual sensing, as well as maintaining resolution, the number of cameras must be scaled in proportion to the distance squared IN ADDITION to the similar scaling of the resolution. The accuracy of angular alignment of each camera also becomes more critical.



Natural evolution faced the same quandary...

And solved it the same way twice...

Ophiocoma wendtii

cephalopods

vertebrates

To understand more specifically how cocodex uniquely relaxes subsystem specifications, consider the visual sensing and rendering of the face...

As was argued previously, a plain video feed doesn't work. There are three primary strategies for visual sensing and reconstruction of the human face for tele-immersion that have been shown to do better:

a) Image based
B) Low parameter avatar
c) Volumetric

Mature tele-immersion will probably synthesize all three.

The visual subsystem of a tele-immersion system must:

- Make plausible demands of bandwidth and latency from network services.
- Render the face in a way that works for users.
- Be able to work with available or soon to be available transducers.

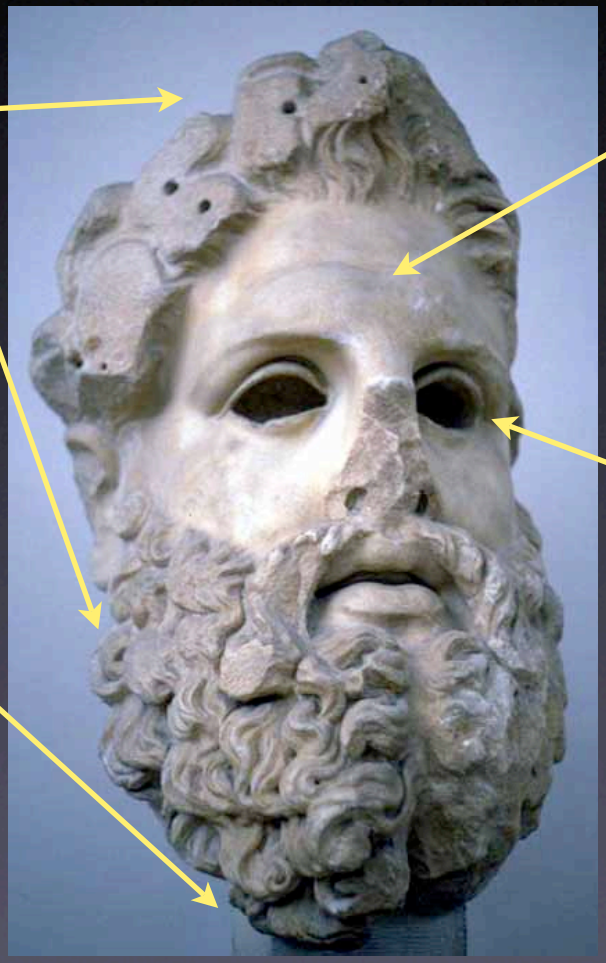
These three requirements turn out to be deeply related.

All three visual tele-immersion strategies (a) Image based, B) Low parameter avatar, c) Volumetric) as they are now understood are highly sensitive to the face moving out of frame, off-axis head pose, and changes in illumination.

Cocodex has an advantage over previous face rendering platforms in that all the cameras, lighting, and display elements are kept within ideal ranges of orientations and positions relative to the face.

People are so specialized at facial perception that facial representation presents unique challenges.

Hair is tricky. Since it's the primary object of human design on most heads, you have to get it right. But it's too detailed to render perfectly. You can't make assumptions about limits to its shape or extent (me: guilty as charged.)



Skin has never been rendered realistically by any means, digital or not, so you have to find a flattering punting strategy. (Renderings can still be distinguished in "blind" tests.)

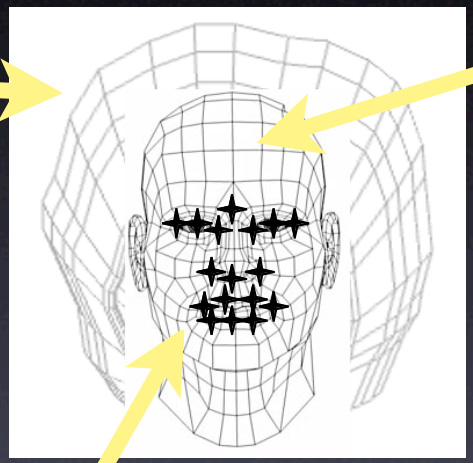
The eyeballs have perhaps come closer to capture by photography, though stone was poor at the job. Minute details in the skin surrounding the eyes seem to be extremely important.

Head pose contributes to communication, and all previous head renderings of any kind had control of it in the service of esthetics.

Zeus, in his heyday

"Compound Portraiture," the face sensing and reconstruction subsystem of cocodex, will blend all three tele-immersion techniques...

Volumetric and/or hull-based approximate method to capture instantaneous shape of serendipitous "big hair", hats, jewelry, etc. Rendering is blended into transparency at periphery, creating a volume halo-like effect in the worst case of giant hair or hat.



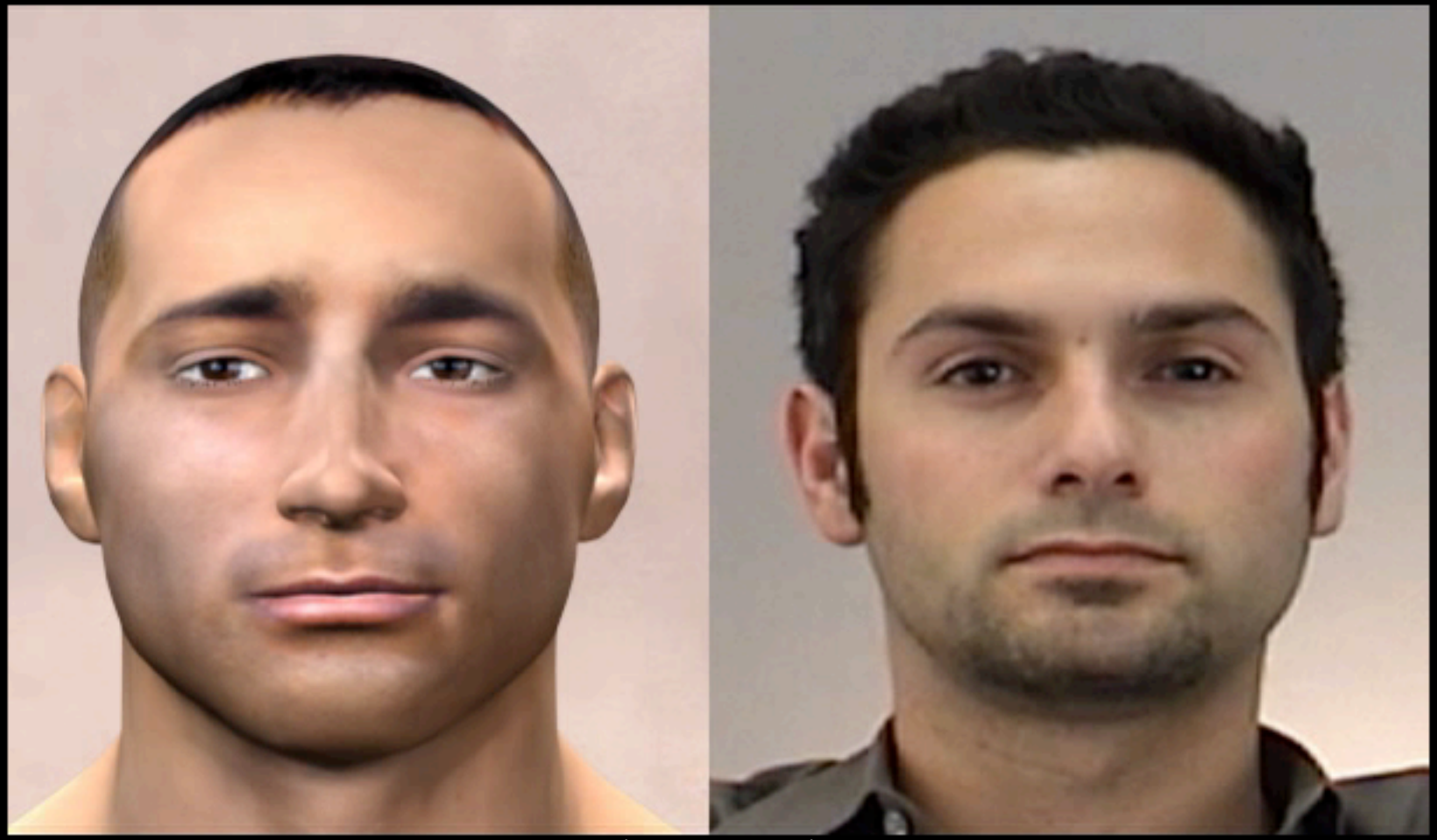
Skin made of surface texture blended from all cameras. For each millimeter region on the surface of the face, the closest and most parallel camera is the primary source of texture. Since cameras are positioned at a selection of angles, the resolution of facial texture doesn't degrade off axis.

Assuming "Big Bertha"/9 Lenticular display (explained below!) and best known cameras, cocodex generally will be able to calculate sub-pixel accuracy in facial textures. Because facial landmarks are tracked, actual transmitted resolution will probably vary according to the importance of the zone of the face in order to reduce latency. Corners of the eyes, for instance would always be rendered at maximum resolution.

Some esthetic filtering/lighting will probably be applied to skin to repeat the esthetic punting strategy that has kept art going through the millenia.

Low parameter avatar tracking points: reliable over wide range of motions because of constancy of relative camera position and Bayesian data fusion, and subject to predictive latency reduction. This data changes facial pose (shape) and aspects of motion in skin texture.

Here is an example of low parameter avatar tracking...



(This is a movie)

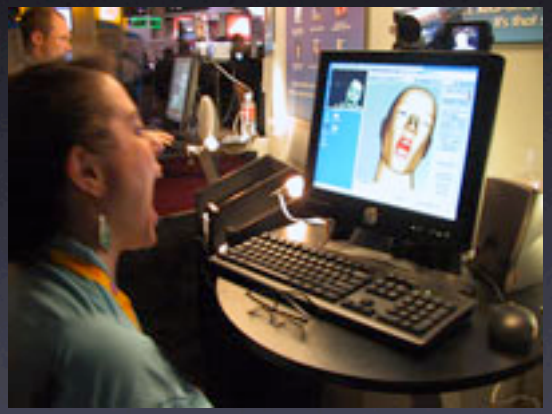
Eyematic demo

Credit to: Christoph von der Marsburg, Hartmut Neven, Ulrich Buddemeier

Jaron was Chief Scientist of the company that made this demo...

Film clip was of 60 tracking points at 60hz on a 1G PC in 2002- BUT we could **only** achieve this performance if the face stayed in the frame, the lighting was very consistent, and the face didn't turn too far from looking straight ahead (degradation would begin at about 25° off axis.) This is STILL the case, even though the algorithms have improved in many ways.

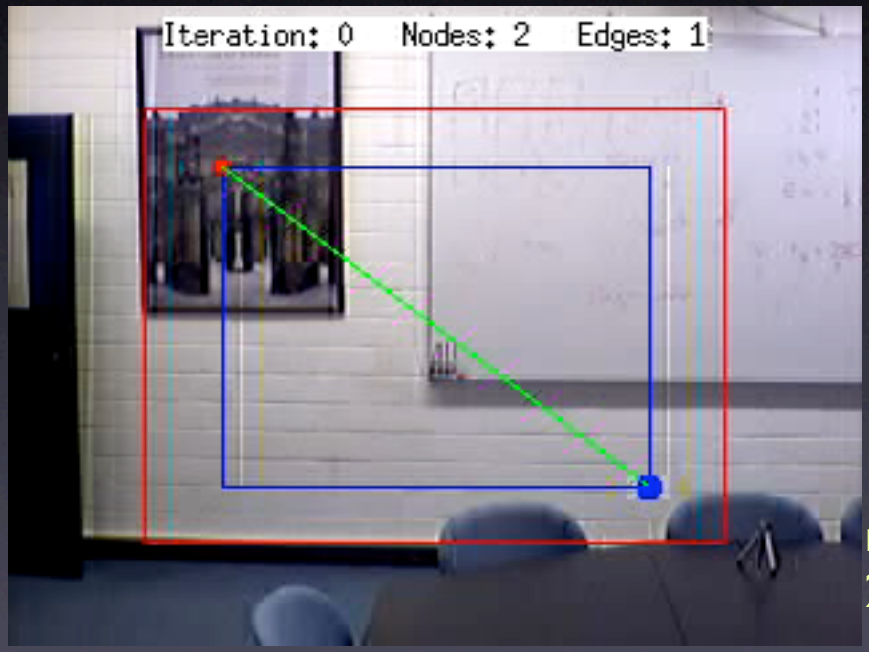
Some of the tracking points used for low parameter avatar control- note these are not physically present on the face!



Of course machine vision will improve, but moving from continuous tracking to discontinuous tracking will be a big leap and progress will be unpredictable.

Tracking is accomplished by a two step process: finding wavelet jets and graph matching to a face prototype.

Texture map for "shape-only" low parameter avatar.



(This is a movie)

ref: Malsburg et al, USC, 2003

Here's how it looks if you just find wavelet jets.

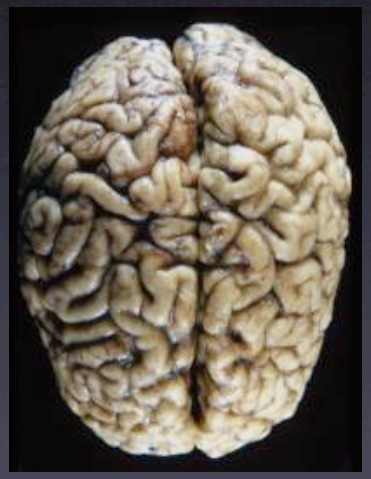
This type of machine vision is important because of the problem of latency.

Latency between coasts of photons through fiber

≈

Latency between ears of signals through neurons and synapses

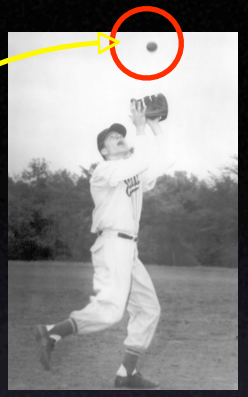
>50ms between coasts, given photon speed of 2/3C through fiber, non-geodesic routes, etc.



>15ms from auditory event to corpus callosum

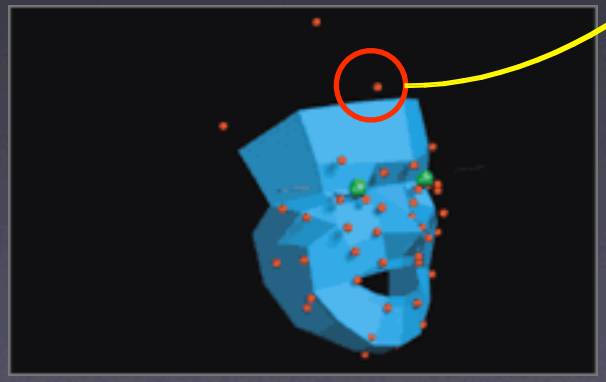
Both would appear to take too long.

The way to reduce the damage done by latency, both in natural and artificial systems, is prediction. The brain is constantly predicting where parts of one's body are about to be, as well as the bodies of others and other events.



Prediction isn't perfect, but can work well enough. Some of the least predictable human motions, like eye saccades, are not quickly or well perceived anyway, and this is no coincidence.

Predicting facial pose control points becomes like predicting where the baseball will be...

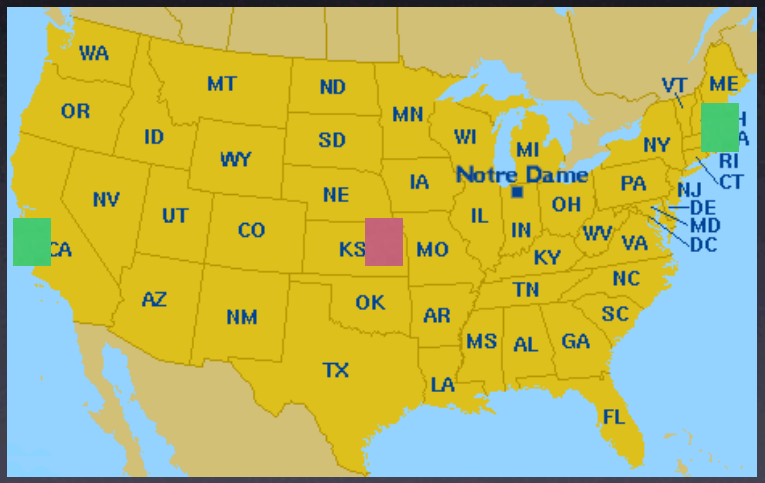


Low parameter representations of people (or anything else) are best suited to the predictive techniques we know about. That's why the avatar method has a future as one aspect of rendering in tele-immersion.

At Internet2 we learned that tele-immersion will need a geographically-sensitive infrastructure...

Tele-immersion servers will need the most recent data possible (to improve predictions,) so now is the time to buy land in Kansas!

Tele-i station in Berkeley

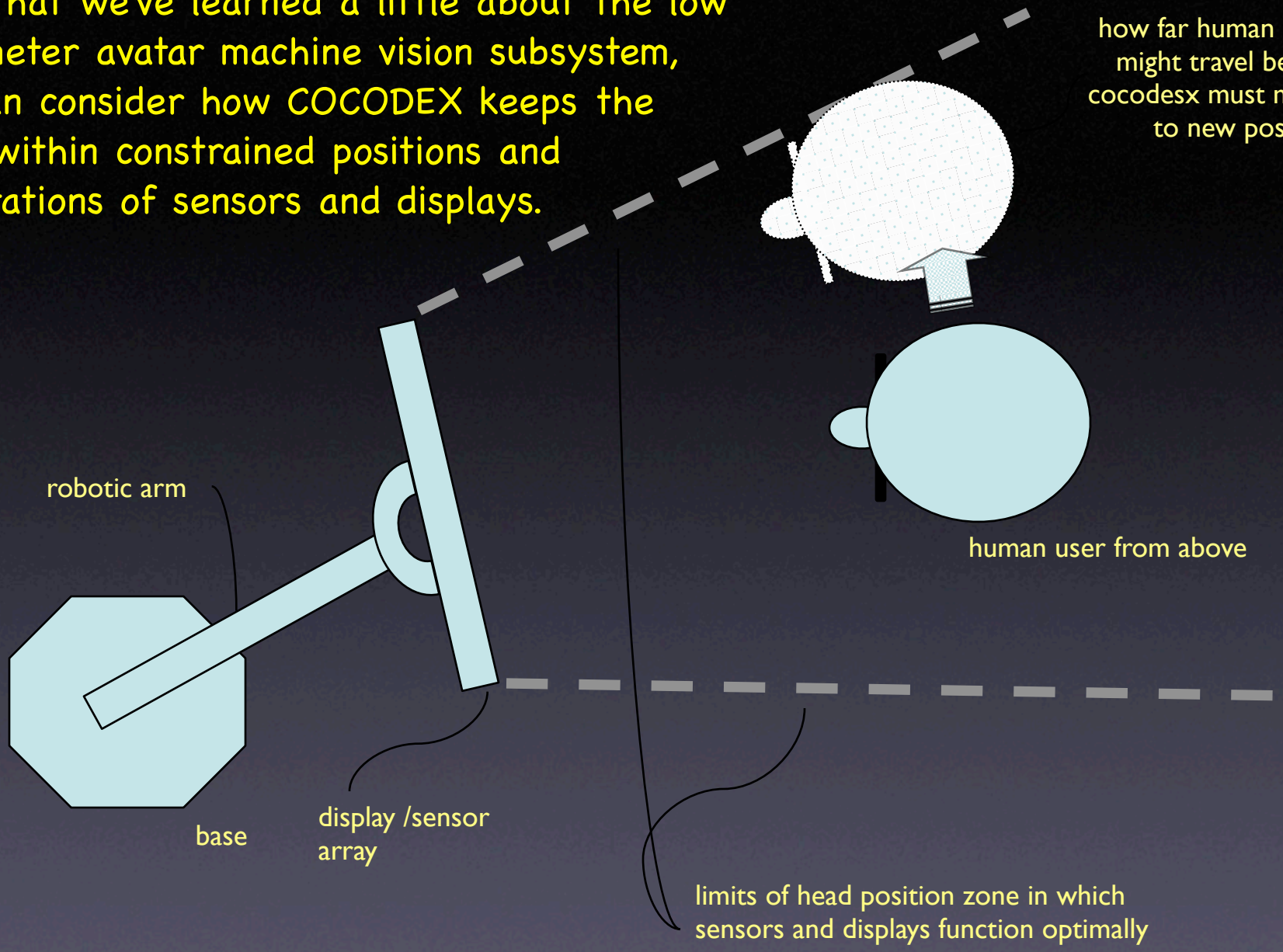


Tele-i station at MIT

Tele-i server

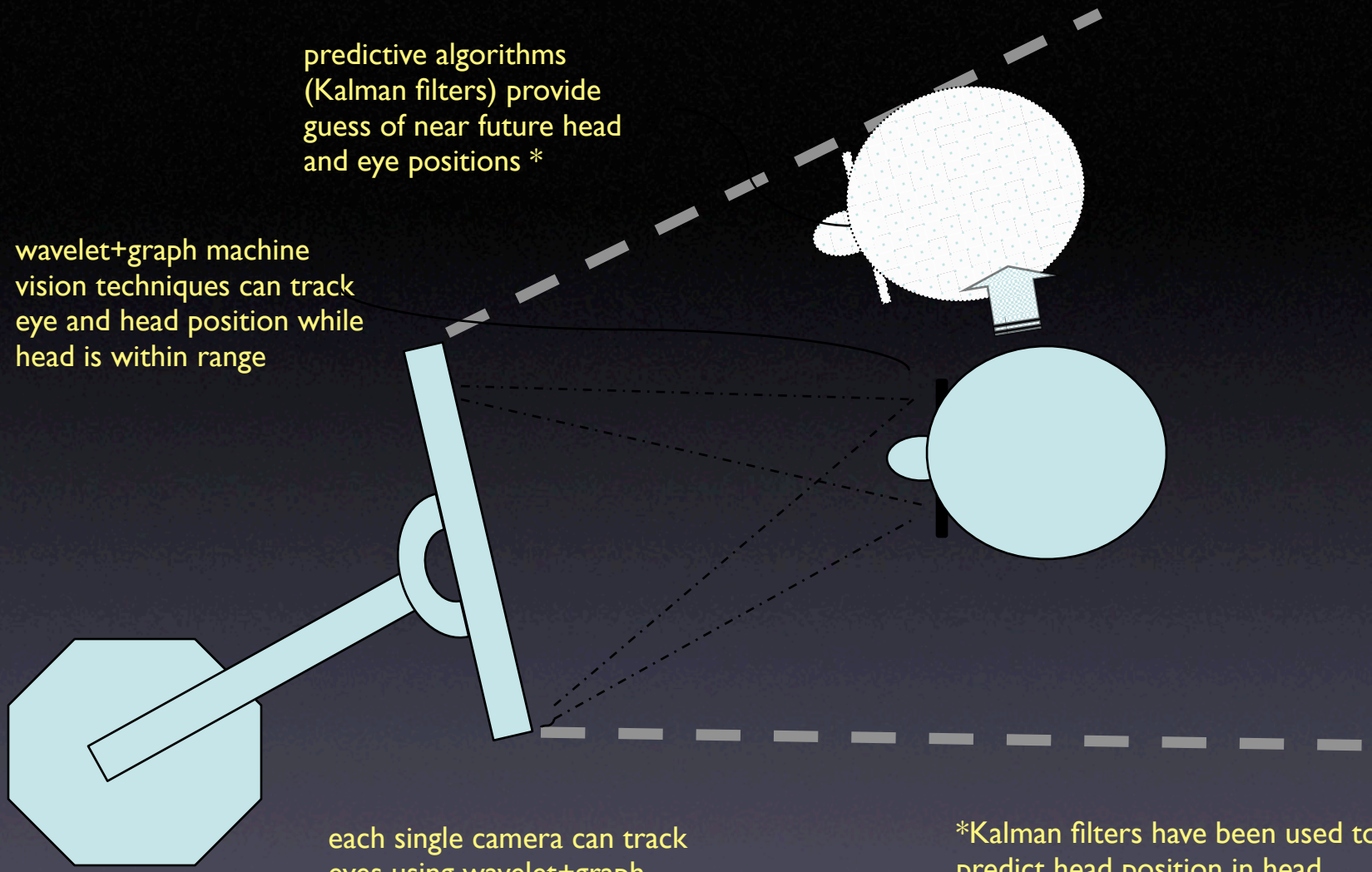
Low parameter prediction might be able to make the visual channel about as fast as the audio channel. Maybe audio will even be partially predicted someday, perhaps based on visual analysis of pre-sonic mouth motion (already partially demonstrated!)

Now that we've learned a little about the low parameter avatar machine vision subsystem, we can consider how COCODEX keeps the user within constrained positions and orientations of sensors and displays.



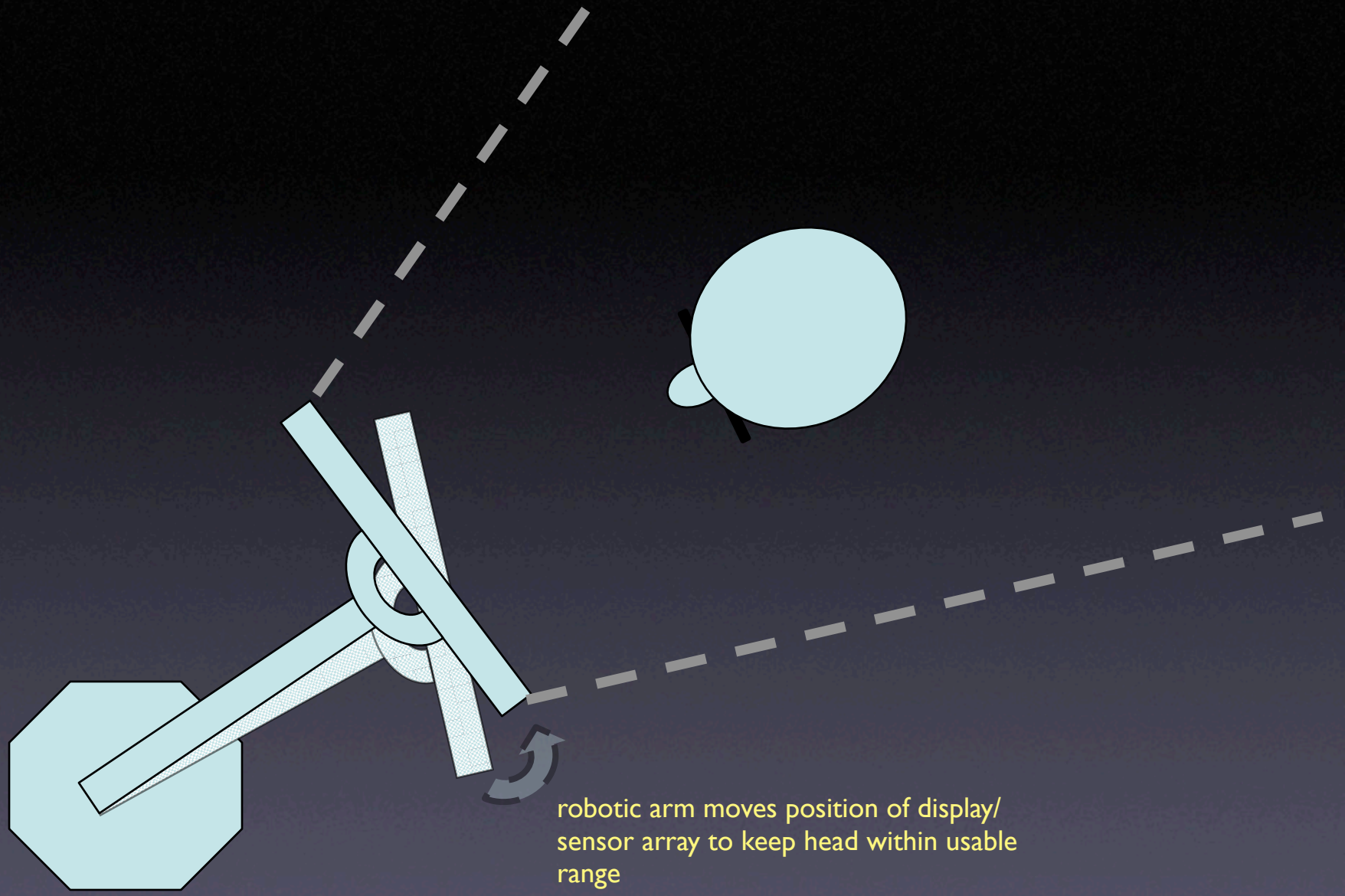
predictive algorithms (Kalman filters) provide guess of near future head and eye positions *

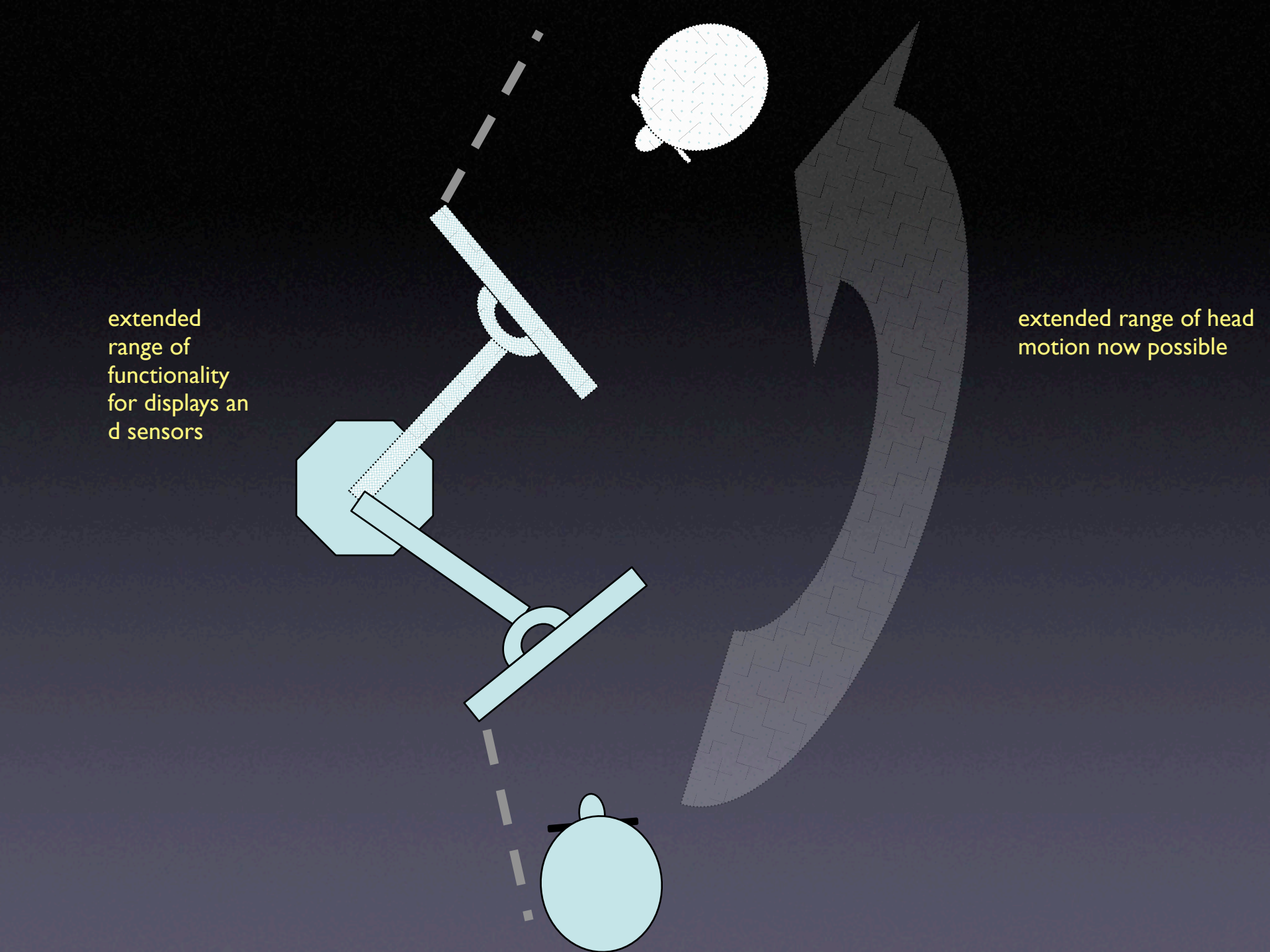
wavelet+graph machine vision techniques can track eye and head position while head is within range



each single camera can track eyes using wavelet+graph method, and Bayesian methods will synthesize an even better guess from whole camera array

*Kalman filters have been used to predict head position in head mounted displays for many years to reduce apparent latency, so similar application is already well understood


robotic arm moves position of display/
sensor array to keep head within usable
range


extended range of functionality for displays an d sensors
extended range of head motion now possible

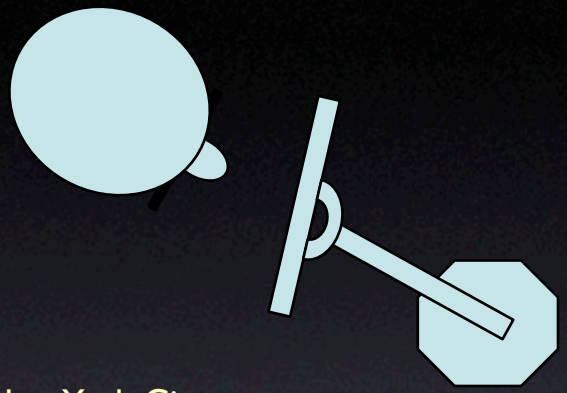
By making the cameras mobile, proven techniques can be applied to make machine vision robust enough to allow cameras to serve as feedback devices for guiding their own mobility.

Note: Machine vision serves as the sole head tracker for COCODEX!

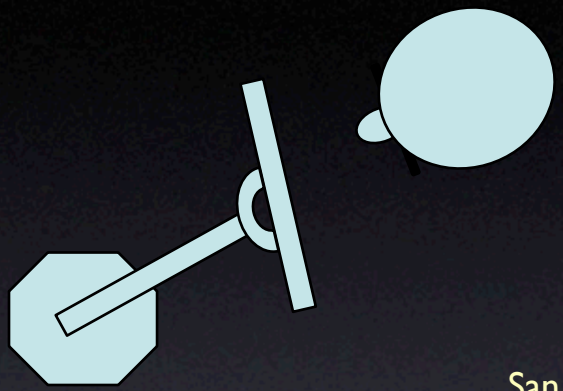
Variations of this principle are true not only for visual sensing, but for illumination, audio sensing, audio display, and autostereo visual display.

But before considering those issues, let's recall why we care...

>2 users can enjoy normal range of motion and FULL DUPLEX!

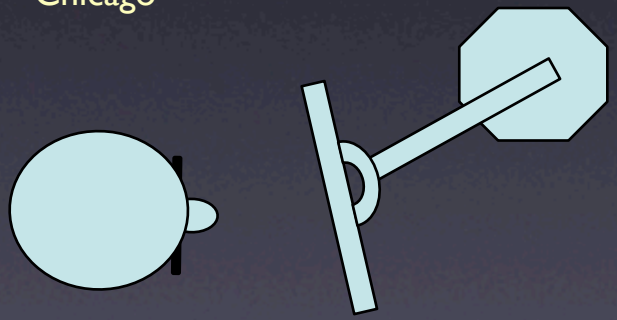


New York City

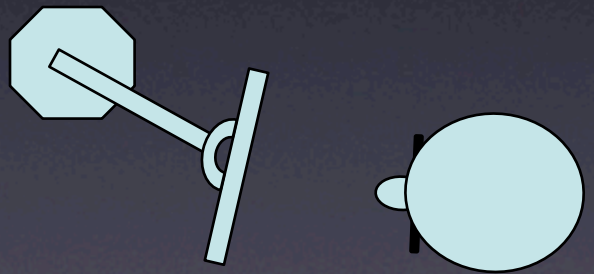


San Francisco

Chicago

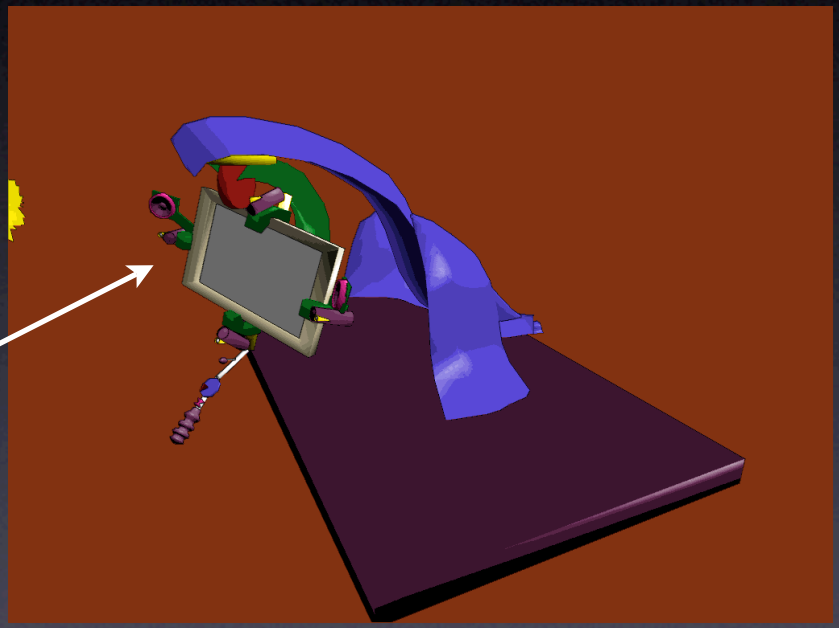


Boston



With true lines of sight!

So the argument is that COCODEX makes known components robust enough to work and also solves the full duplex problem.



"Brontosaurus" variant; bigger but faster.

Aren't robotic arms expensive?

expensive

Frequently they are, but in part relative to accuracy.

The cocodex arm needs to know where it is with great accuracy, but it doesn't have to get itself to a targeted position with much accuracy. It only has to keep the head within approximately constrained positions and orientations. Cocodex can be a sloppy mover!

cheap

Of course it'll be important to keep the cocodex i/o "payload" weight as light as possible. Fortunately, display, camera, lighting, and other transducer technologies are all trending downward in weight.

The only two missing pieces that would have made cocodex impossible until recently were the cpu power for machine vision and LED light weight displays.

Could cocodex wallop you?

Utterly crucial question!

As will be explained below, cocodex is designed to be moved by external touch as well as on its own power and will not be intrinsically resistant. There's also a very large research community studying collision avoidance. Even so, in a liability-driven time, wallop prevention will have to be rigorous.

Further wallop prevention and management:

1) "Zero weight" design.

2) Light payload.

3) Cameras and other sensors on back and on base as well as on front to support collision avoidance.

4) Padding, just in case.

5) No sharp edges, just in case.

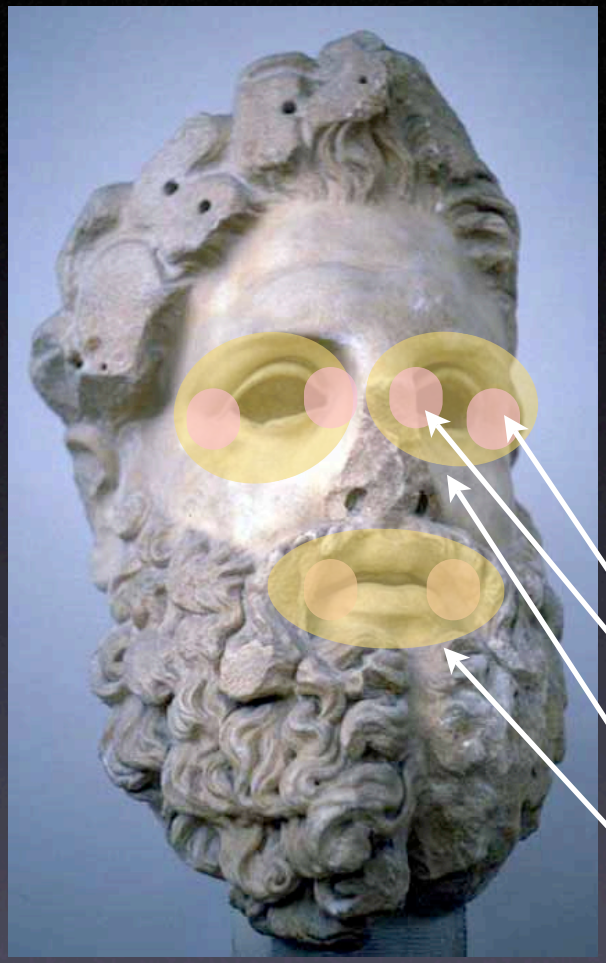
My plea is that the potential benefits justify the next phase of research, and that safety can be addressed in later phases.

What about bandwidth?

Purist low-parameter avatar tracking only updates the facial pose, but not textures on the face. To update the whole texture would require bandwidth similar to a video stream.

Compound portraiture allows for the selection of small areas of interest which are transmitted at the highest resolutions and lowest latencies.

Other textures on the face are updated with less resolution and more slowly.



Notice that Zeus loses expressive power with these small areas obscured...

Blending, fading, lighting, and shading algorithms are essential to combining out of sync contributions to changes in facial texture. Primary motion in face will still come from low parameter avatar morphing of geometry underlying texture.

Image-based techniques will optimize textures, and animated skin textures aligned to synthesized viewing positions should create wonderful rendering of skin transparency.

Small areas of greatest importance.

Medium-sized areas of medium importance.

In earlier experiments, we've found that people sometimes don't want to be rendered as realistically as possible.

You'll have access to a virtual mirror to be aware of personal appearance and you'll be able to make adjustments.

Someday, cocodex will probably include a low-parameter editor for skin tone, feature sharpness, fidget filtering, wardrobe change, etc...




We've already added hats, hairstyles, and whatnot to people.

Status of hypotheses in the argument for the viability of compound portraiture:

On this first point I claim good intuition based on decades of work...

- The prediction is made that compound portraiture will render people well enough, where plain video feeds and individual tele-immersion techniques have not. STATUS: There have been combinations of two tele-immersion techniques at a time, but not all three at once. This is a crucial area where implementation and testing is needed.

- Compound portraiture should support latency reduction through use of predictive filtering on low parameter avatar subsystem. STATUS: This claim has only been tested indirectly, in that similar techniques have been applied to HMD tracking over networks.

- Compound portraiture should reduce bandwidth needs by varying resolution according to area of interest and reducing requirements for low latency update of many areas of the face. STATUS: This idea has been implemented and tested in various way by various groups and appears to work.

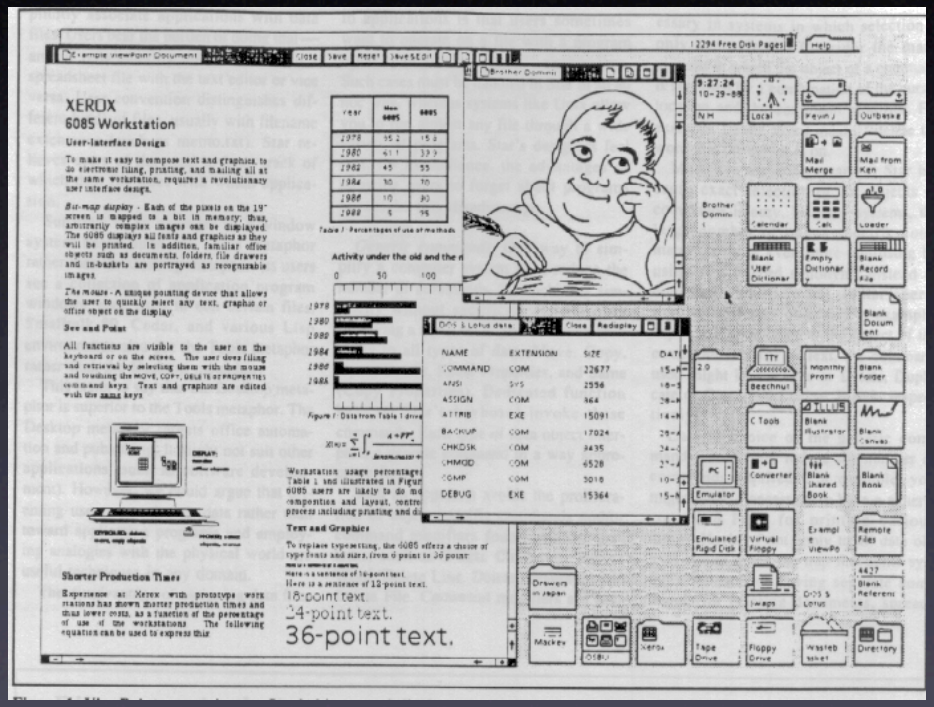
- The subsystems of compound portraiture only work if the face, lighting, and cameras remain within tolerances of relative position and orientation. STATUS: Must implement and test cocodex arm. No other methods of acquiring adequate sensing data have been articulated at this time.

Part Four:
The "Special Room" Problem; Why it's important and how COCODEX addresses it.

All the proposed instrumentation strategies for visual tele-communication (except COCODEX) that solve even a subset of the basic usability problems require special rooms.

If Carolina Cruz-Neira had been a grad student at Stanford instead of U Ill, she might not have invented the CAVE for her dissertation, because of the expense and the politics of finding that special room.

Immersion isn't the only driver of the special room problem; An even deeper issue is the screen real estate crunch.



Viewpoint OS, late 1970s era software, originally introduced on "Star" workstation.

The principal solution thus far has been the use of scrolling and overlapping windows, which were first developed at Xerox PARC.

Unfortunately, the overlapping windows solution is having a hard time keeping up with patterns of use.

Montreal Police Command Center

Hi res display wall at Princeton.

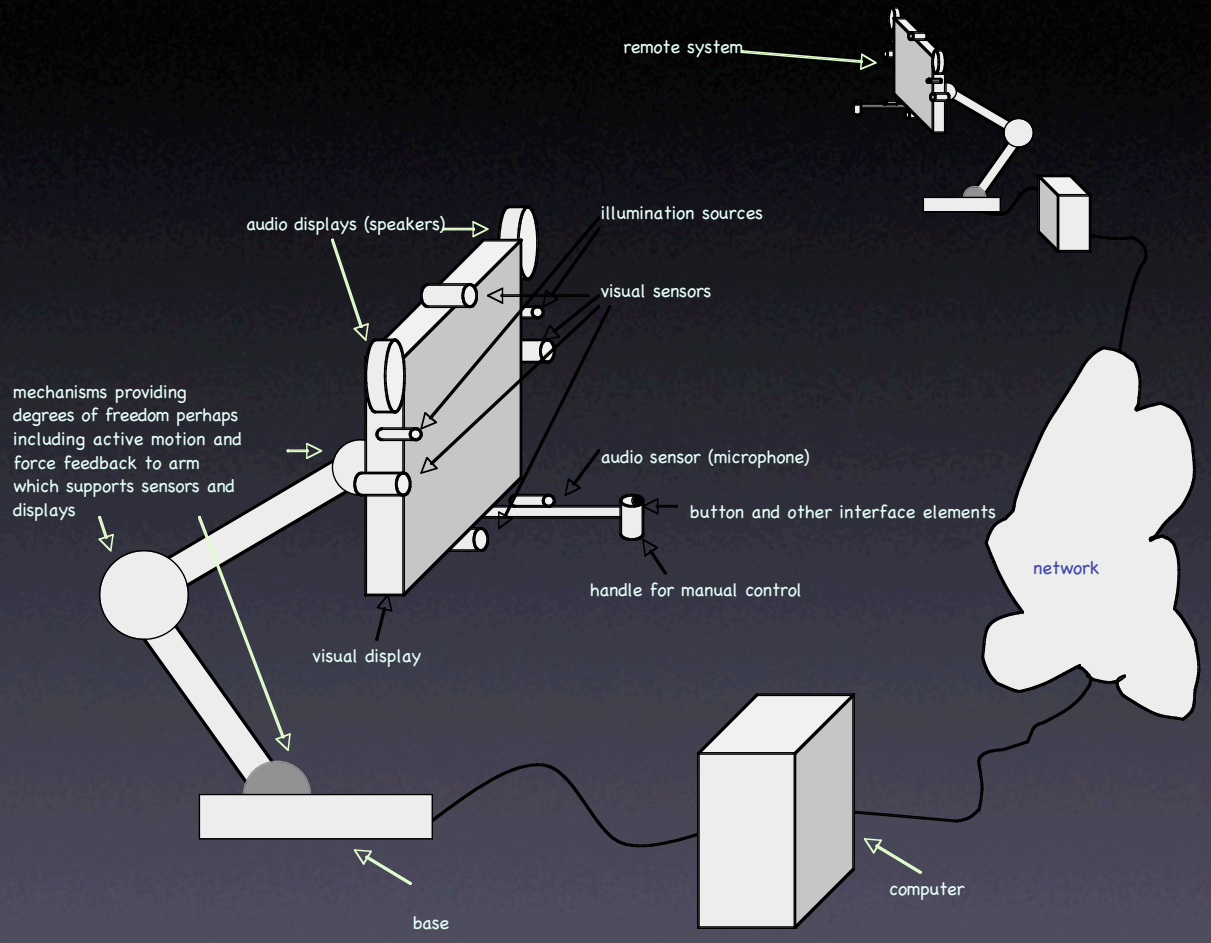
Behind the scenes at a typical hi res display wall.

Special rooms are increasingly needed to convey multiple simultaneous streams of information or high resolution visuals.

It's not just the expense and the politics- it's also that special rooms demand a break in life flow and workflow. Inevitably entering a special room means leaving other tools behind, such as one's "conventional" computer.

COCODEX has the potential to provide access to hi res images, multiple streams of visual information, AND most of the benefits of immersion without requiring a special room or excluding other devices or patterns of behavior.

Of course one could design a "hair-dryer-like" surround display for cocodex that would provide peripheral vision, but the flat display design would be much cheaper, and the possibility of having workable tele-immersion and many of the benefits of VR while still being connected to other activities in a conventional setting is appealing.

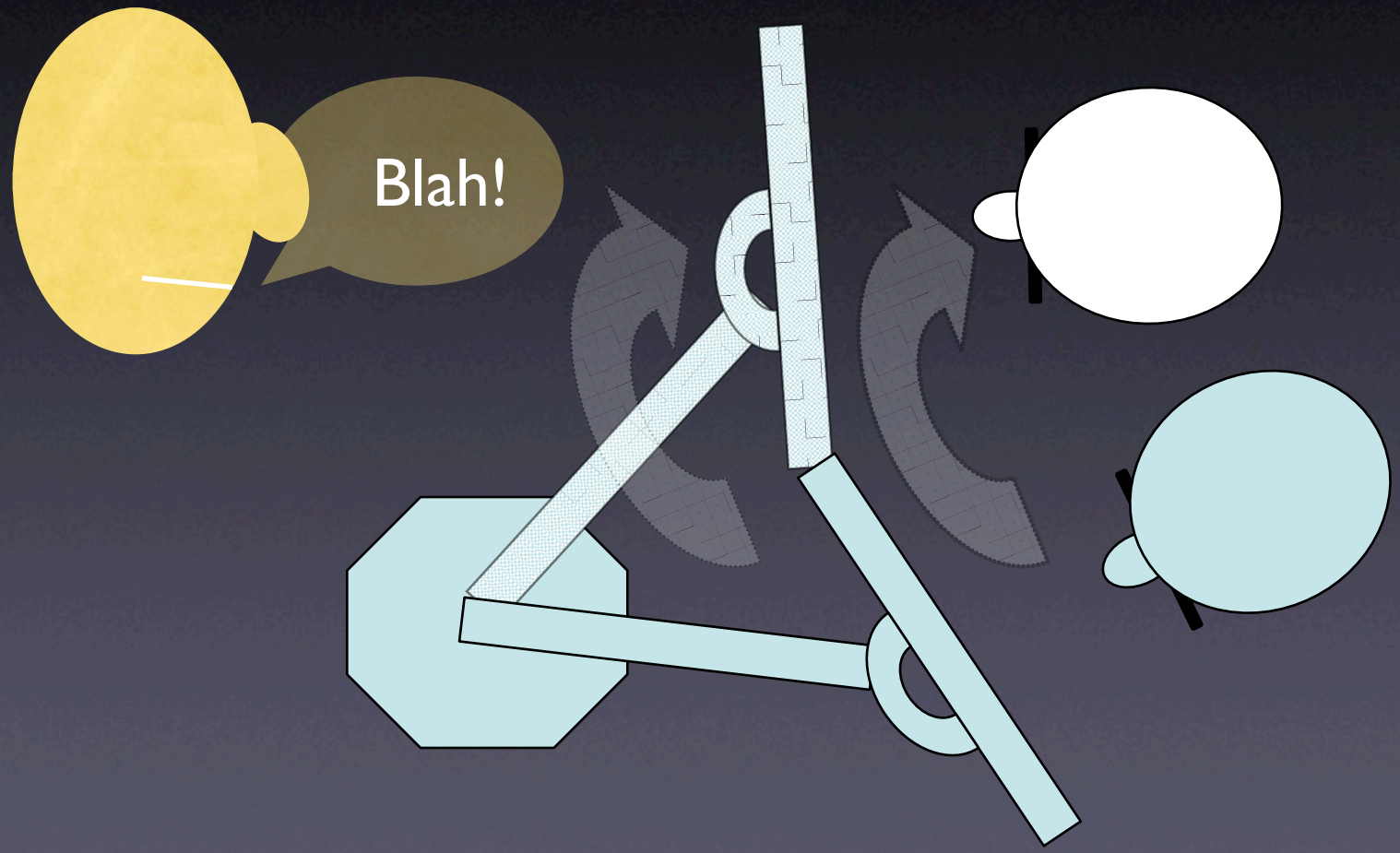


To understand how this compromise (giving up peripheral vision) would work, consider the audio channel...

Stationary speakers are poor at creating effective 3D soundfields, while headphones are so good at it that someone's always raising ridiculous money with the old binaural haircut demo.

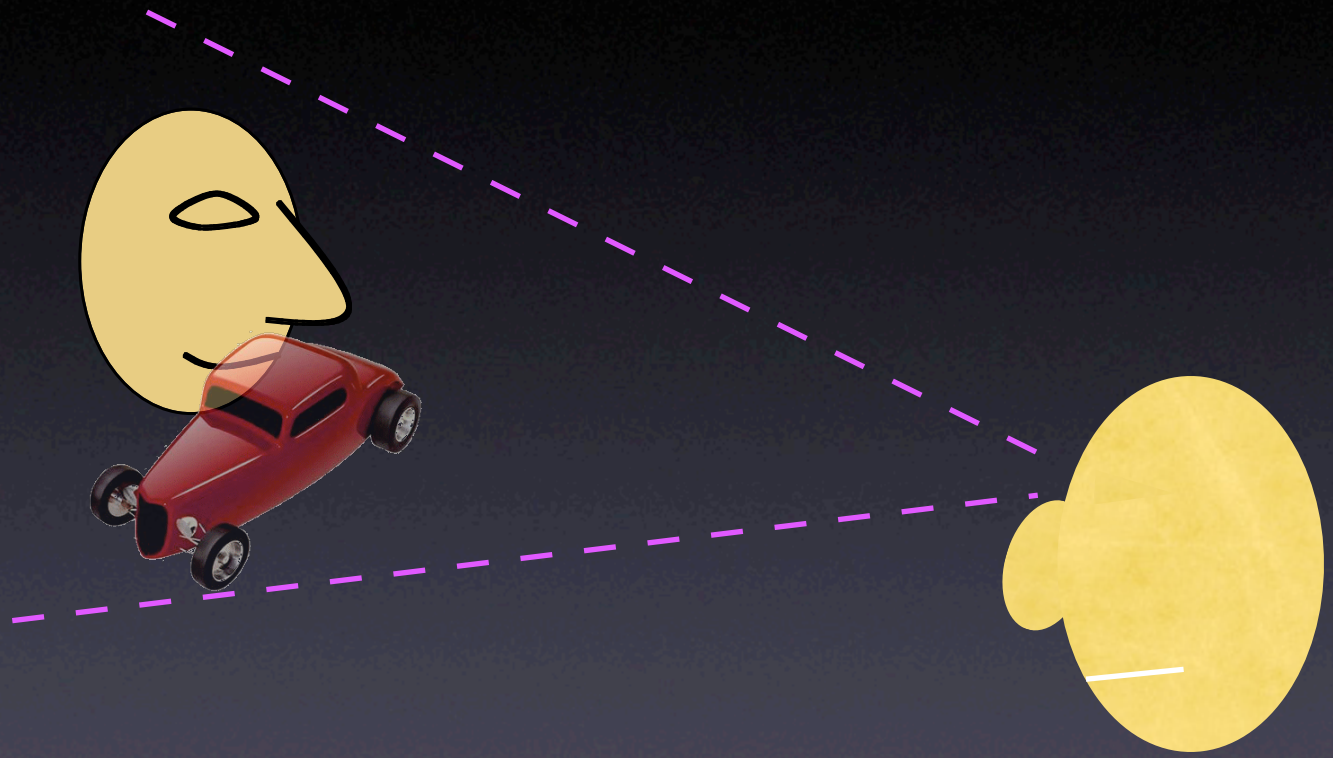
Putting nearfield speakers in motion with cocodex to maintain a constrained relationship with the head should also work. (Demonstrated at about 6"- would need to be extended to about a foot.)

Even though you wouldn't have peripheral vision, you'd have peripheral audio cues, so you'd hear someone to the side and turn to look at him or her.



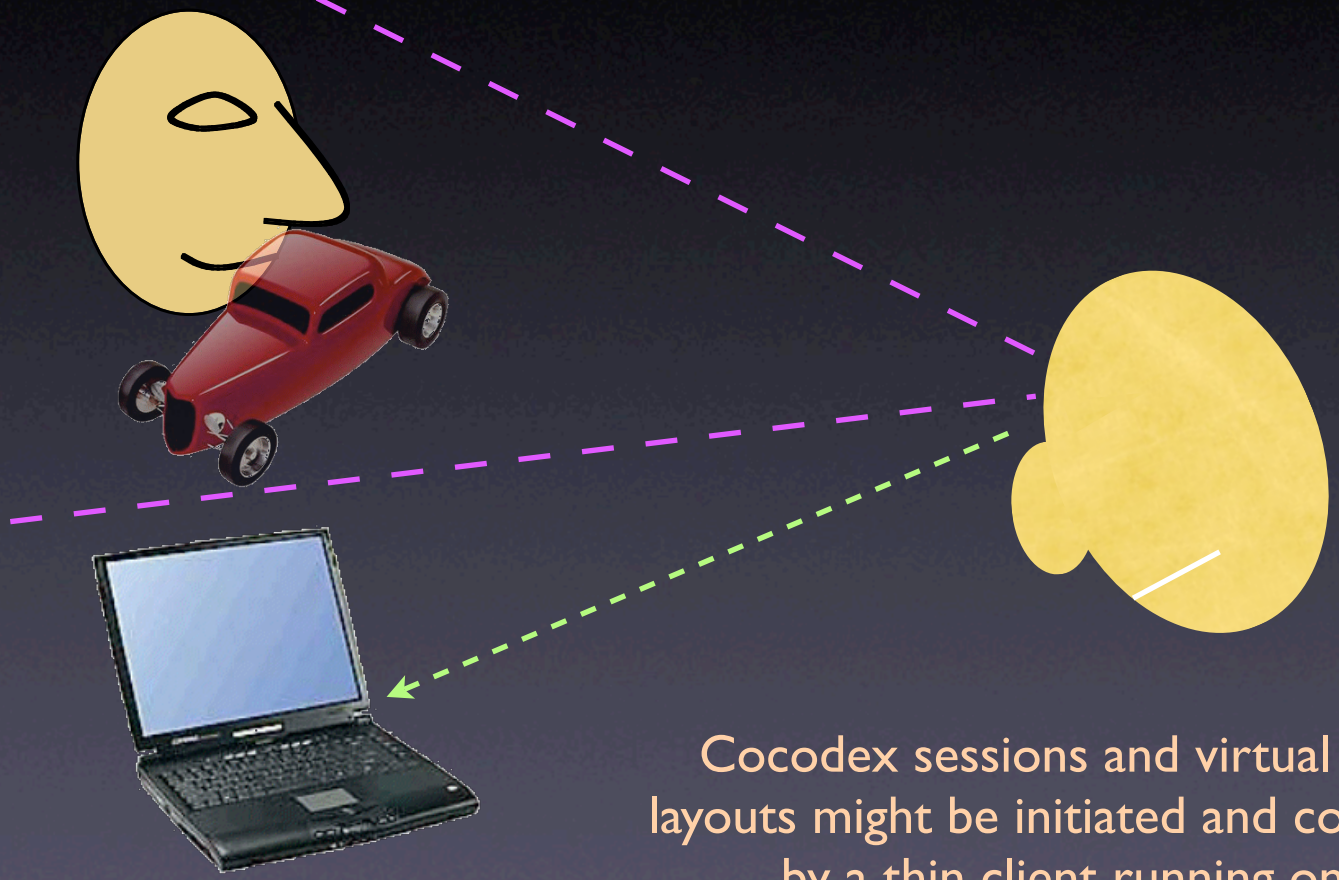
Note that the remote person's position relative to you remains constant (unless that person physically shifts positions.) You can look away from the person just like you do in physical proximity. Looking away is part of communication too!

Cocodex will not always stay in front of your face!



Although the precise rules must be determined through testing, cocodex will probably track you only when you are looking into the "Area of interest" where the remote people and virtual items are located.

With the right tracking rules, cocodex should be able to balance access to the local physical environment, the remote physical environment, and the shared virtual environment.

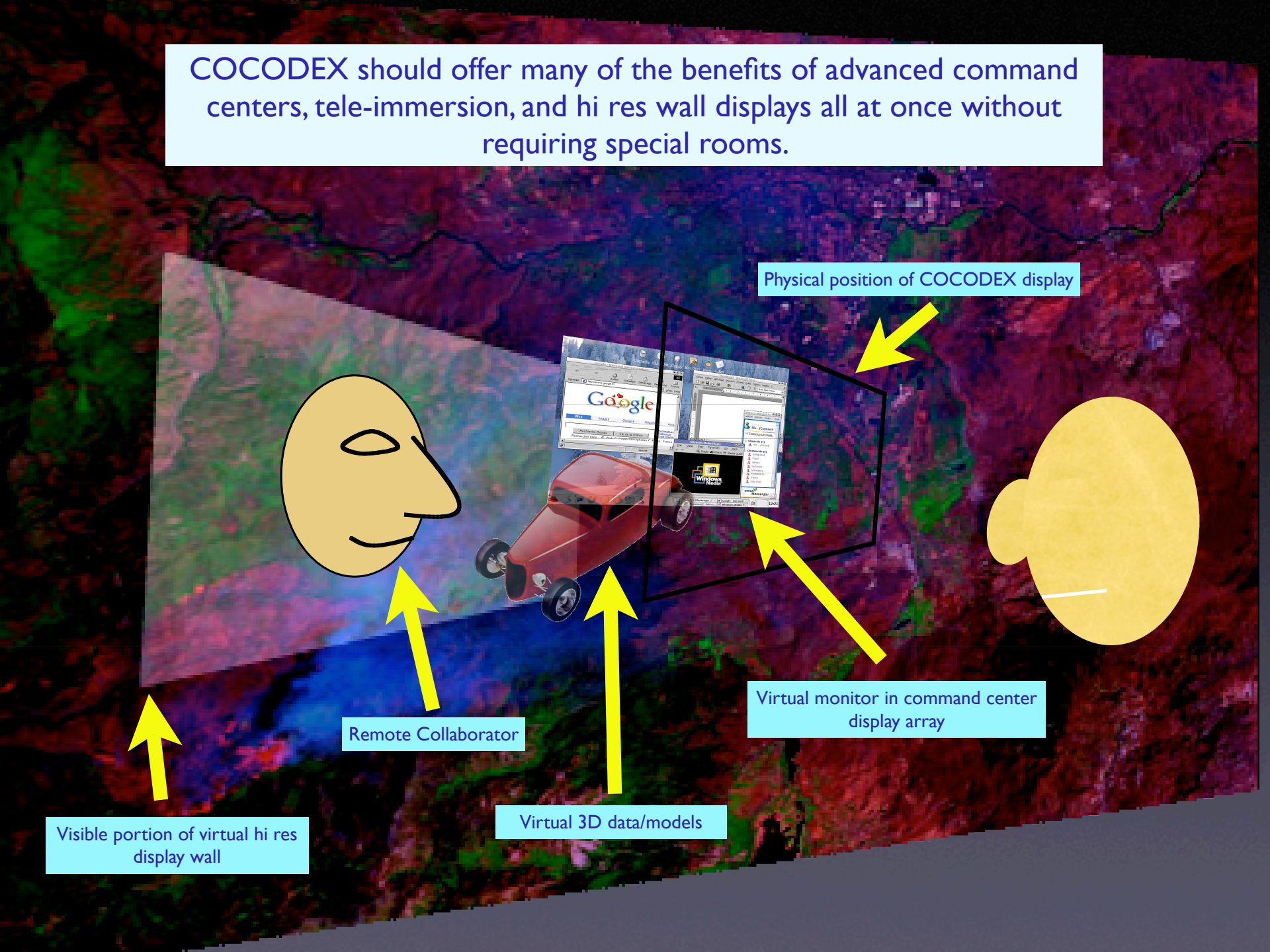


Cocodex sessions and virtual space layouts might be initiated and controlled by a thin client running on a conventional computer.

Since cocodex will only move while it's tracking you, which is when you are paying attention to the world on the other side of it, it is hoped that the motion will not be distracting to the user, and in fact should not even be a prominent part of the user experience.

An analogy is the interior of a car while you are driving. The car is in motion, but not relative to you, the driver!

COCODEX should offer many of the benefits of advanced command centers, tele-immersion, and hi res wall displays all at once without requiring special rooms.

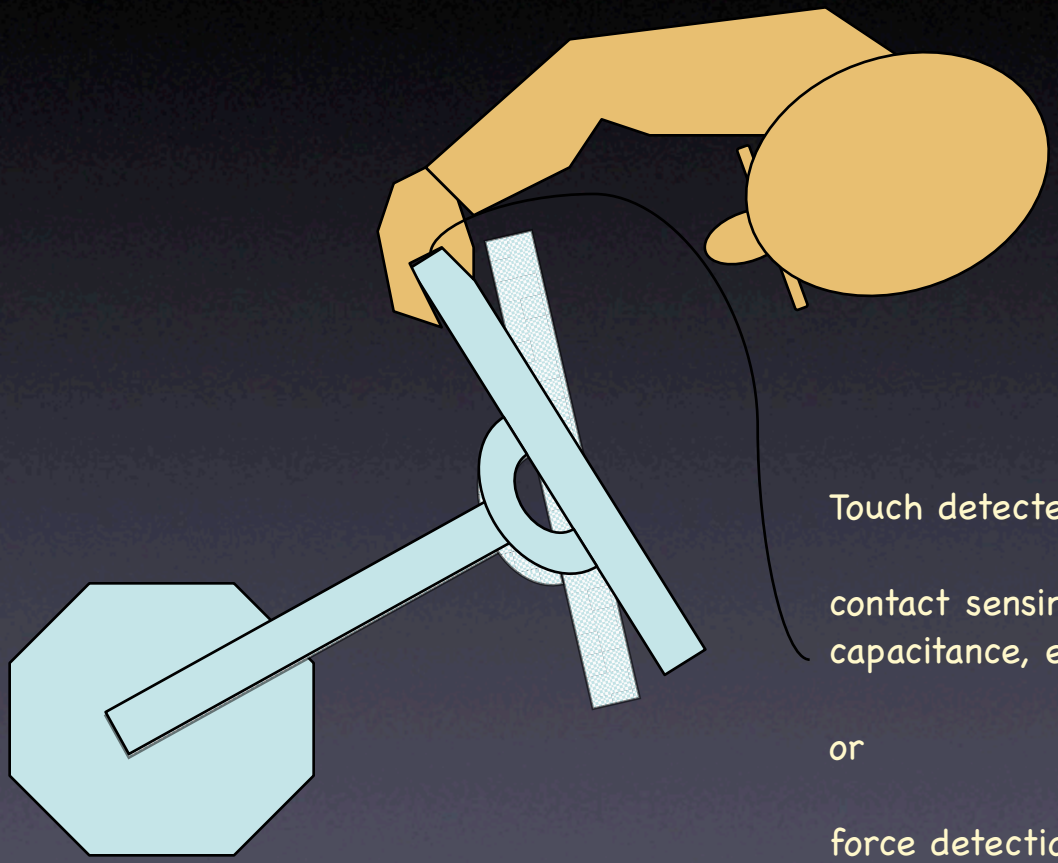


Cocodex might still be distracting to others in the local physical environment, however. This is one of many potential problems that must be researched.

...a special case of the general problem of how people might react to robotic moving objects in the work and home environments.

Part Five:
More about the COCODEX control structure; Cocodex as an improved UI for working with volumetric data

You can also grab COCODEX
and use it as an input device...



Touch detected by:

contact sensing (pressure capacitance, etc.)

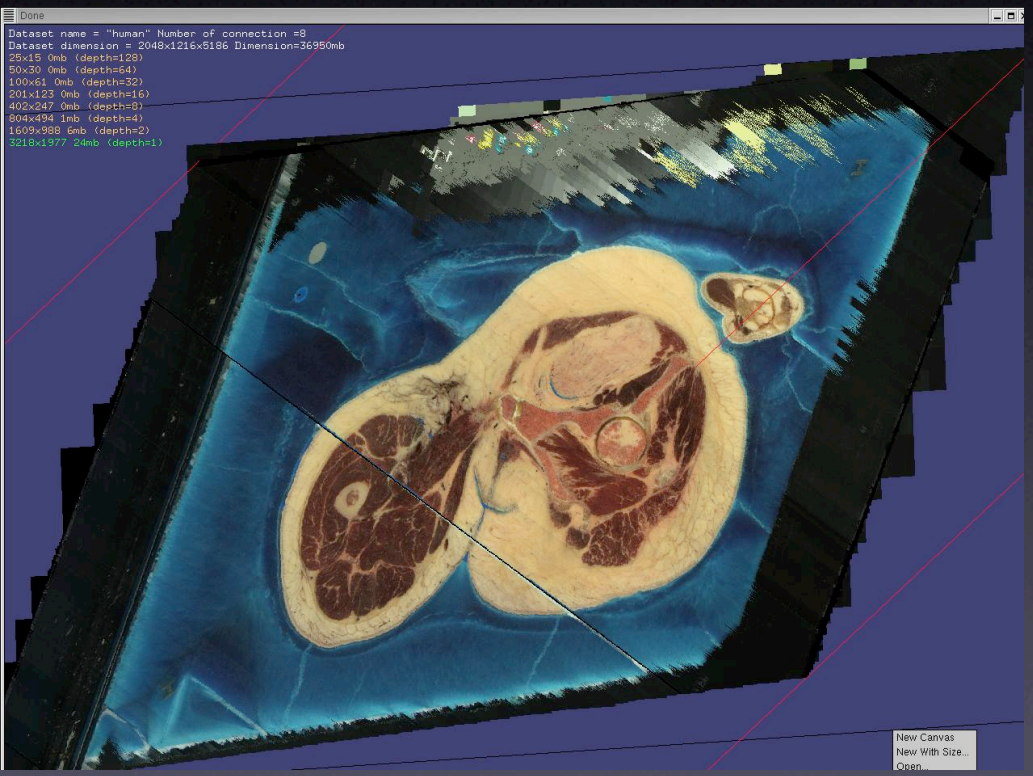
or

force detection

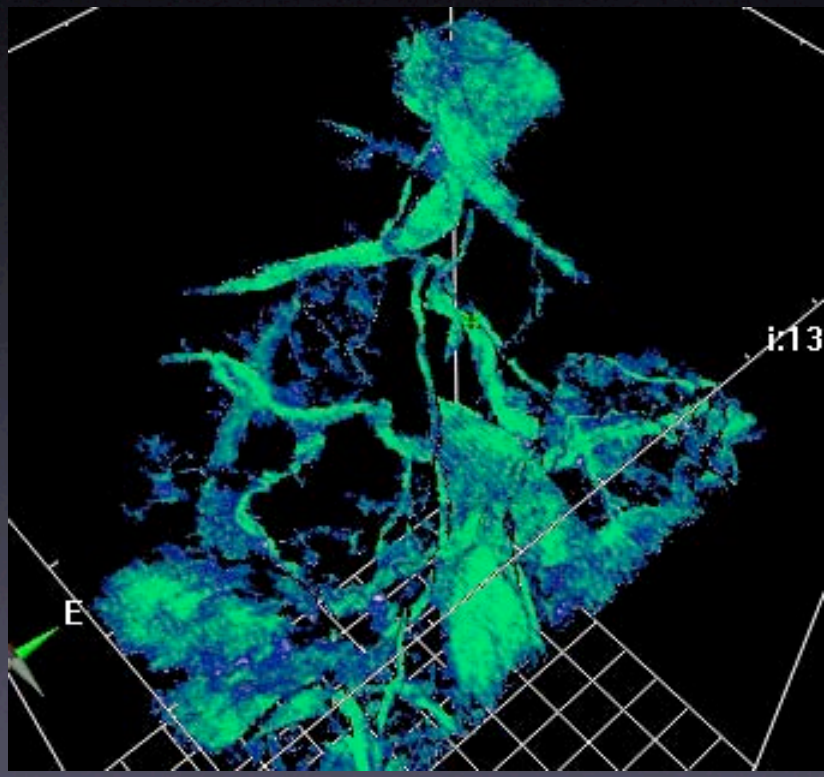
But why would you want to?

Non-rectilinear Volumetric Data is generally hard to Navigate.

But recently it has at least become computationally affordable to do so at interactive speeds…



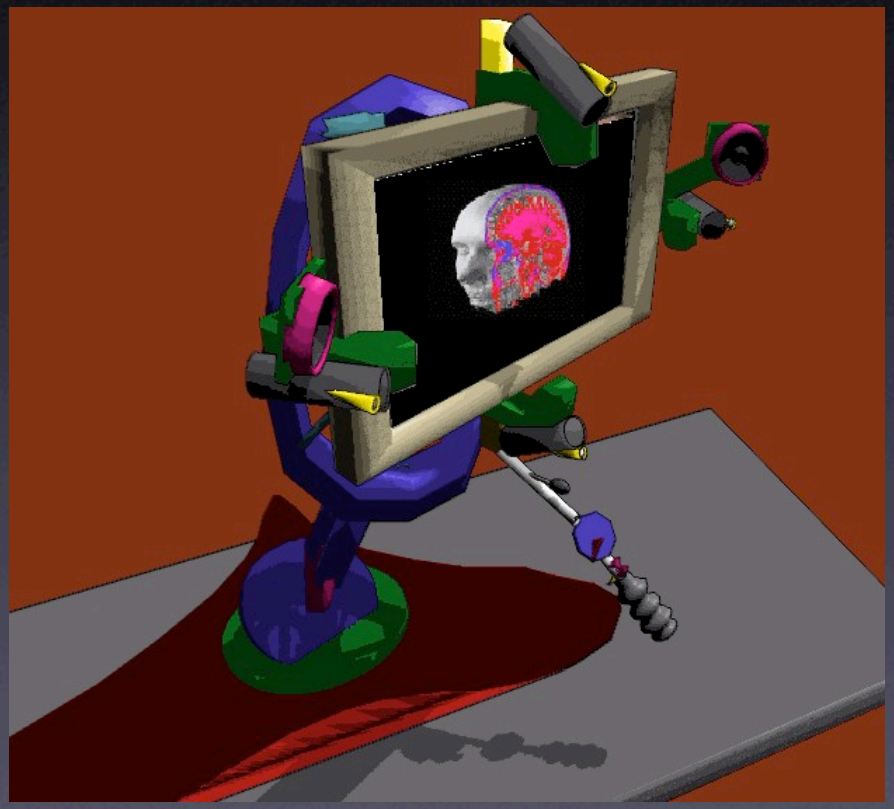
Linsen, Scorzelli, Pascucci, Frank,Hamann, and Joy; UC Davis 2003



GigaViz navigating seismic data on SGI Altix, 2003.

In practice, 6d navigation, especially in dense, non-rectilinear environments, is tough even for seasoned users.

The simple idea is to have the COCODEX position and orientation equal the visual frustum and the cutplane.

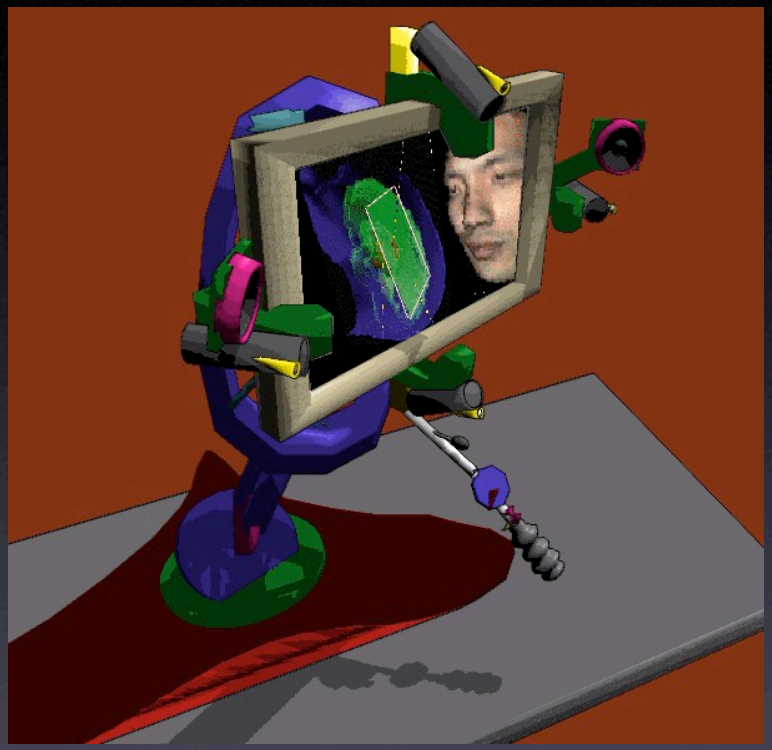


One less mental rotation!

Here's cocodex selecting a brain slice.

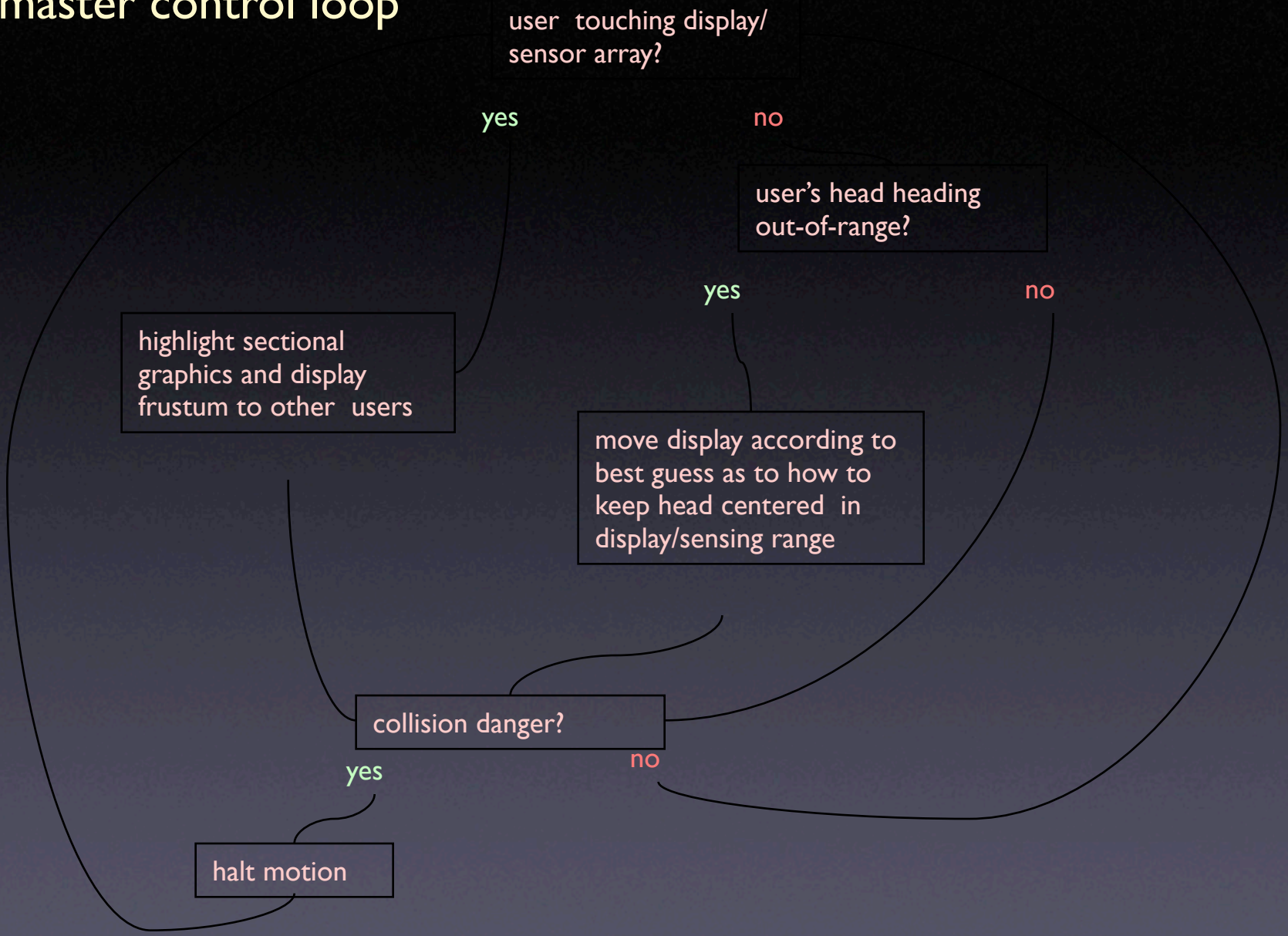
This strategy can also improve collaborative communication between users:

Most important slide in the presentation!

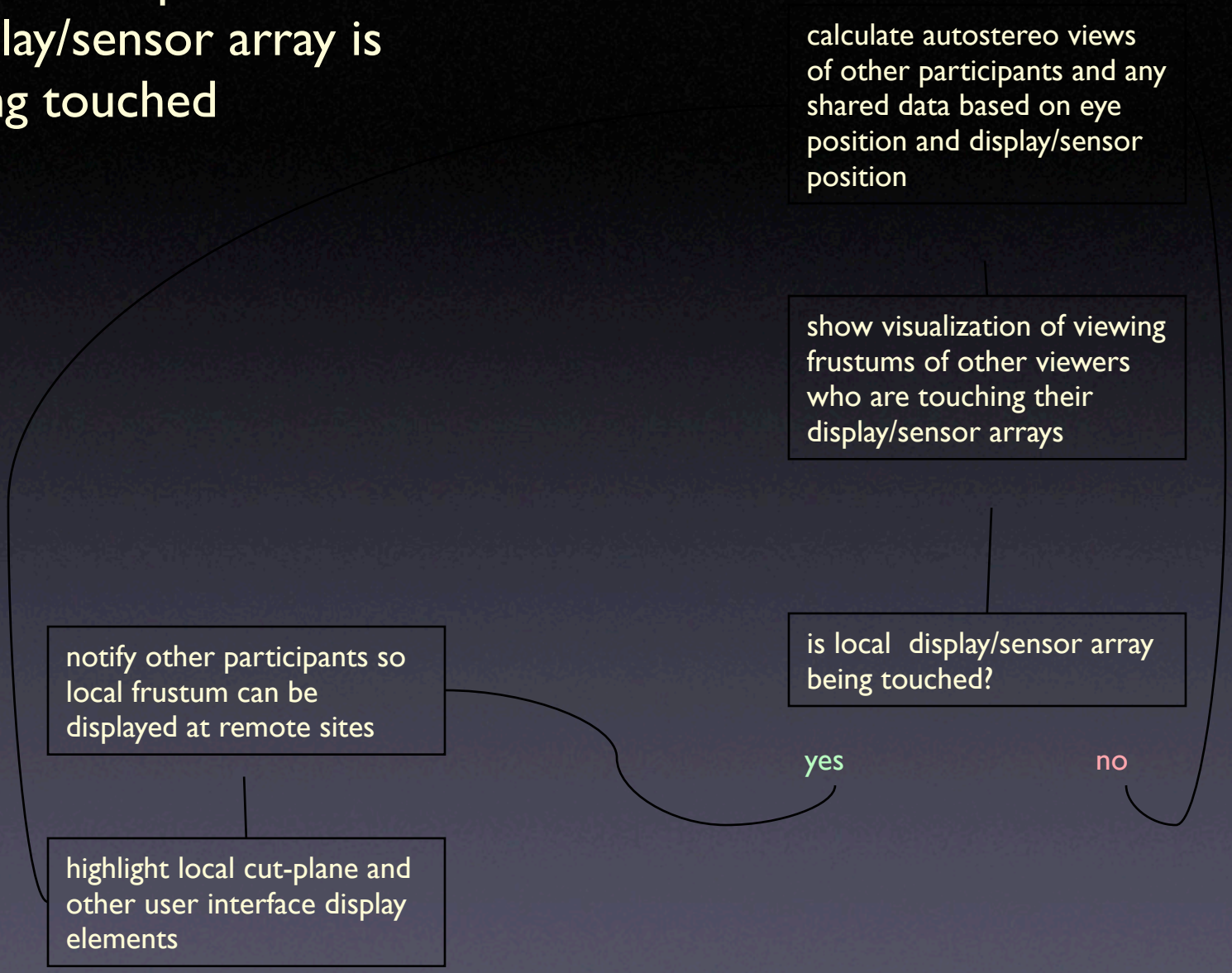


Here we see both the other user's face and the position of that remote user's COCODEX screen. In this way we can tell what the other person is looking at in the shared virtual world.

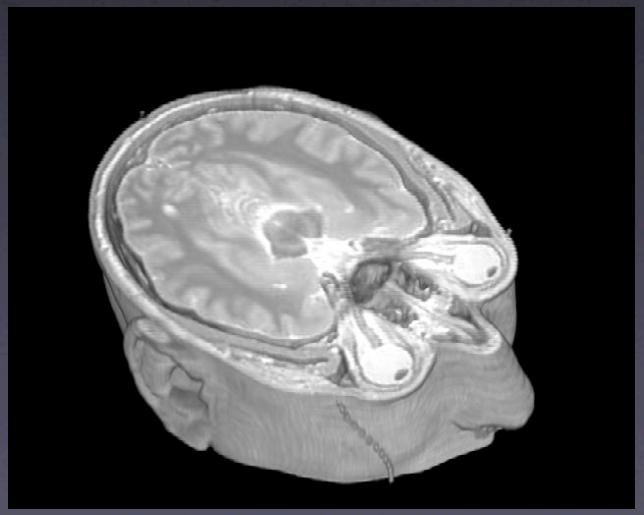
master control loop

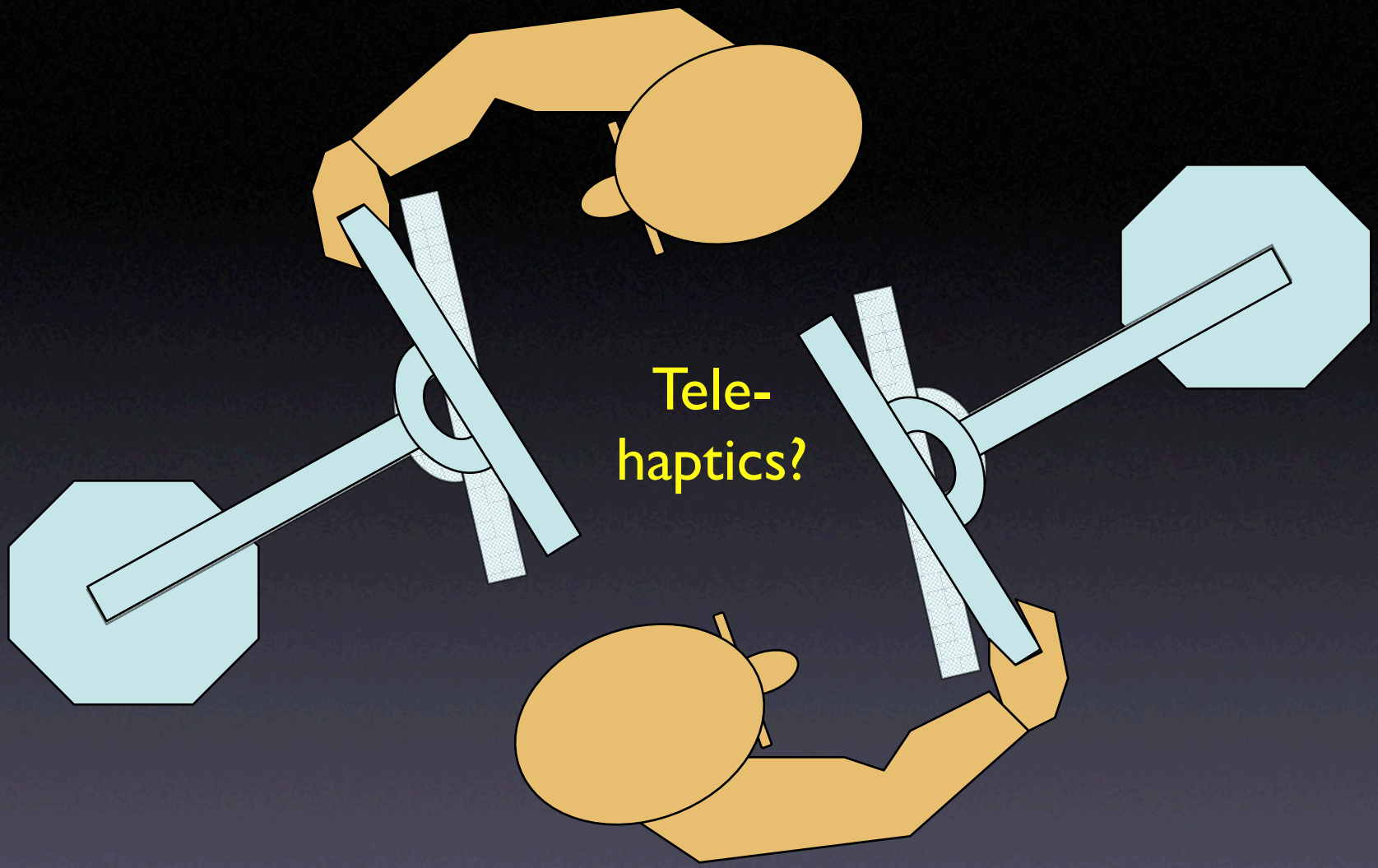


control loop while display/sensor array is being touched



Haptic planar feedback is worth exploring. A plane intersecting a scalar resistance field would be interesting and perhaps more efficient than current point-based haptics. You might get a sense of the shape of a tumor, for instance, faster by planar haptics than by visual methods alone. Curl might also be conveyed. Unity of visual, audio, and haptic exploration through tightly coupled single feedback principle might work well across a wide range of individual cognitive styles.

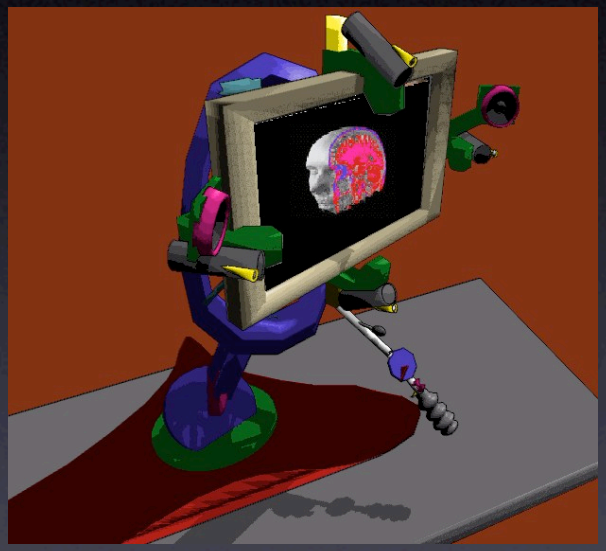




Not only might remote users share a bundle of force and resistance streams in both XYZ and RPY, but this information might be meaningful to the joint exploration of data.

Part Six:
More about the COCODEX control structure; Autostereo considerations

There are two potential types of display for cocodex, diffusive and transmissive. We'll consider the diffusive case first.



I imagine it as a flat screen held by clips so it can be swapped.

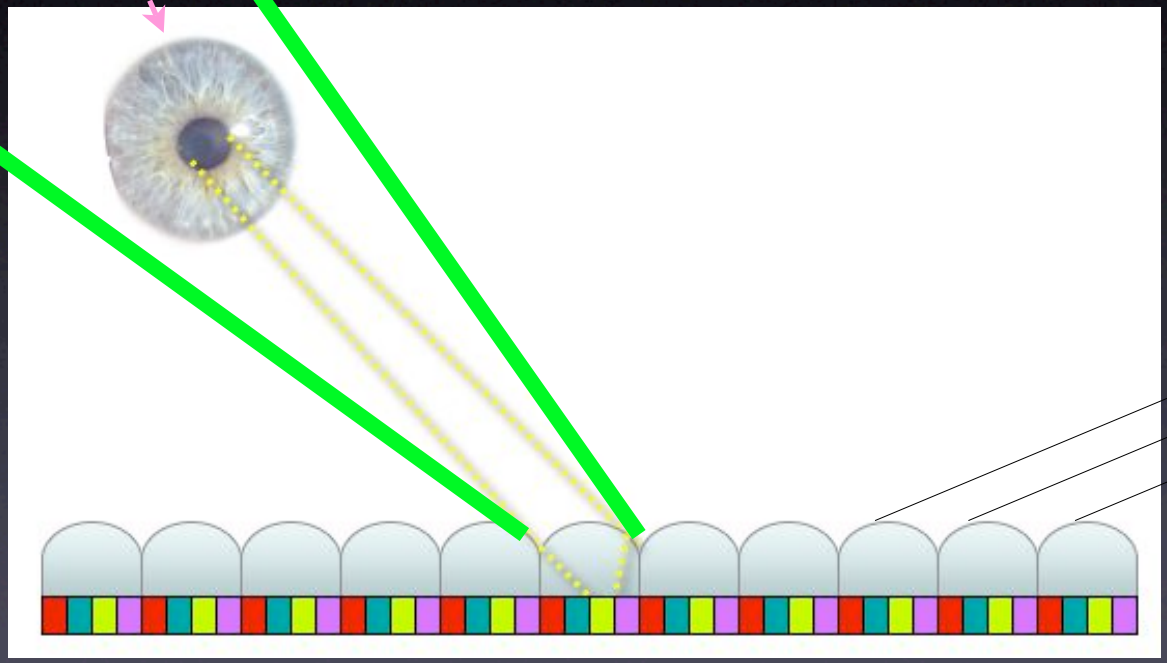
You might sometimes want no autostereo at all, to maximize resolution.

COCODEX can improve lenticular displays and also benefit from them in unusual ways.

Of currently available parts, my favorite is the lenticular version of the "Big Bertha" display from Stereographics. Even with 9 subpixel perspectives, the image has adequate resolution, based on current understanding, and the form factor would yield a pleasant average field of view in COCODEX.

Here's the basic idea of a lenticular display:

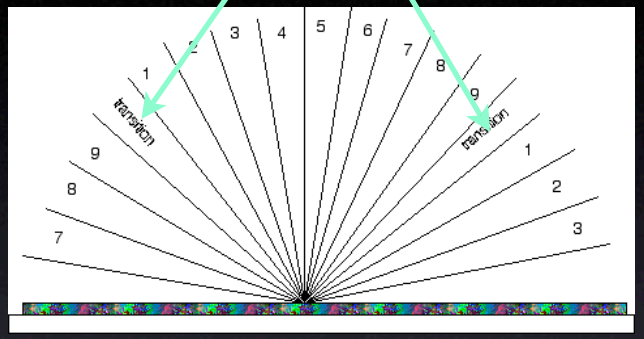
This particular eye sees only the yellow pixel in the direction of the indicated lenticule, but only so long as the eye remains within the green lines. Another eye in a different position will see a different subpixel magnified by the same lenticule.



Lenticules are usually tilted to steal subpixels from both x and y directions in order not to have different resolutions in each axis.

Three cocodex-specific potential improvements to lenticular autostereo:

Pseudscopic eye position wedges to be avoided.



Typical approximate viewing geometry of Synthagram lenticule

1

Autostereo displays place restrictions on head position, so COCODEX might make them usable by people who move around in typical ways.

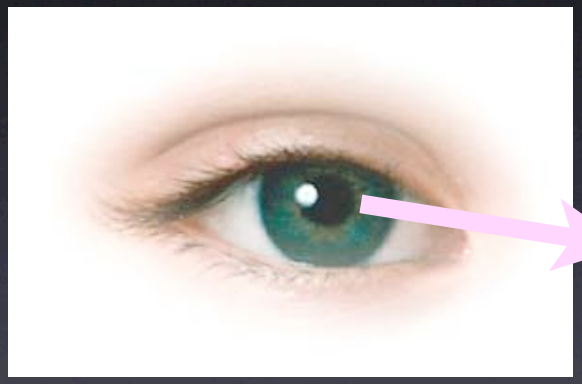
2

Lenticular displays also often suffer from lenticule/subpixel alignment problems: A 3D (volumetric) corrective lookup table could be made for each individual display, but only relative to 3D eye positions, so COCODEX's support of more robust eye tracking makes such correction possible. (Cocodex uniquely enables this potential improvement because eye tracking will probably not be robust enough over the normal range of motion for stationary displays in the foreseeable future.)

Advantage # 3

Eyetracking will allow each eye's perspective to be correct instead of approximate.

Slow return to center when eyetrack data seems poor, or when more than one eye shares a lenticular perspective.



A single lenticular viewing zone.

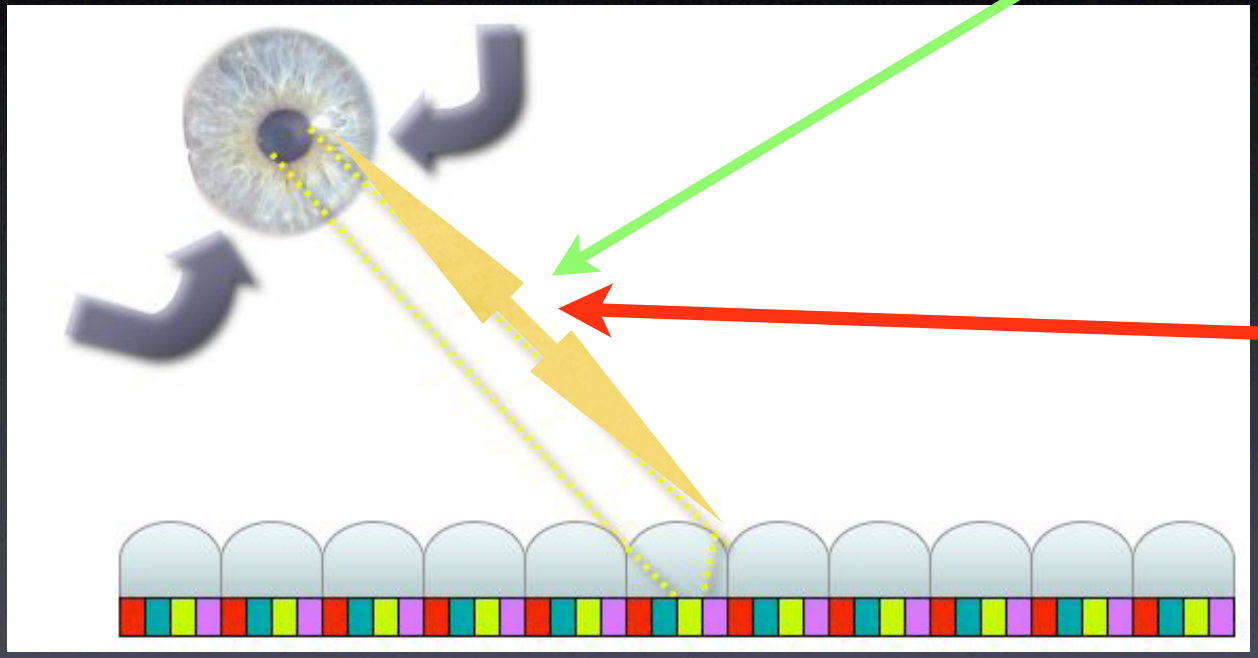
Adjacent viewing wedges will have the same perspective at the moment an eye moves from viewing one into viewing another.

YES, this method might support occasionally gathered multiple viewers on a single COCODEX!

However, there's also a potential problem...

The accommodation distance is an emergent and fluctuating result of the flexing of the human eye's lens in search of a focal distance that brings image edges in the macular zone into sharper focus.

Closer than usual with 3D displays; perhaps 6"- 16" to optimize sensors and field of view, and to keep the "payload" small and lightweight.



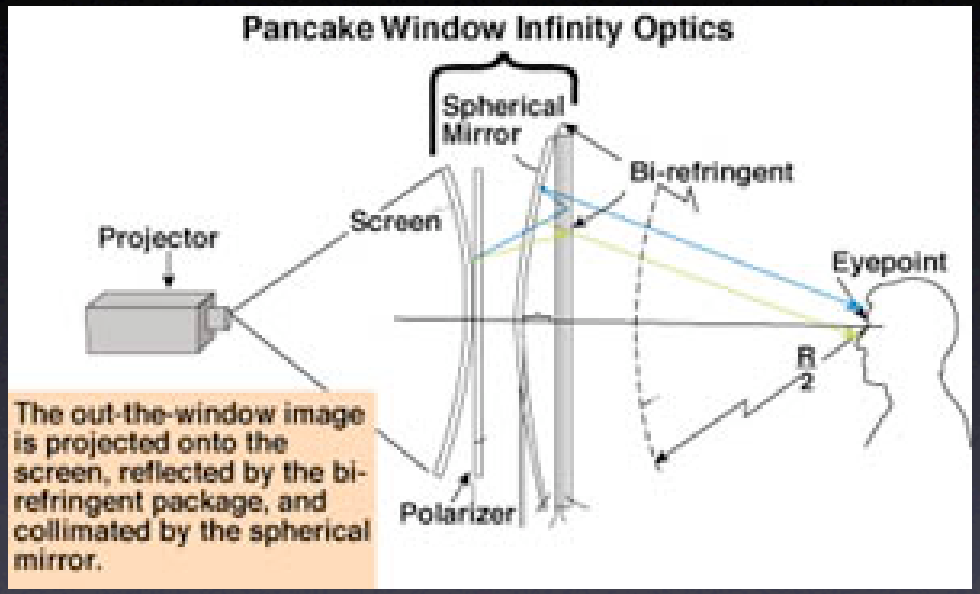
Individual pixels provide no accommodation cues- only edges of pixels, or transitions between pixels provide these cues.

The physical lenticules normally provide the focal distance in lenticular displays since they are crisper than pixelated object boundaries.

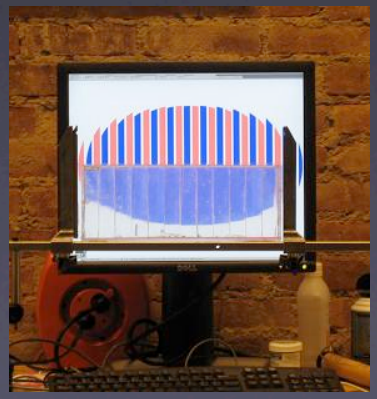
Stereopsis might suggest a different distance.

It's impossible to know in advance whether accommodation will create a usability problem for cocodex. It might not be much of a problem, since lenticular displays confuse the eye as it searches for focused lines by presenting ambiguous and conflicting fine-scale cues (aside from stereopsis.) Since we don't yet know if we have a problem, or what the problem would precisely be like, it's a little early to propose solutions, but here goes anyway...

One idea with some precedence would be to design an infinity optical element with a subpixel selection or masking component.



Warning! More speculative than other ideas in this presentation.

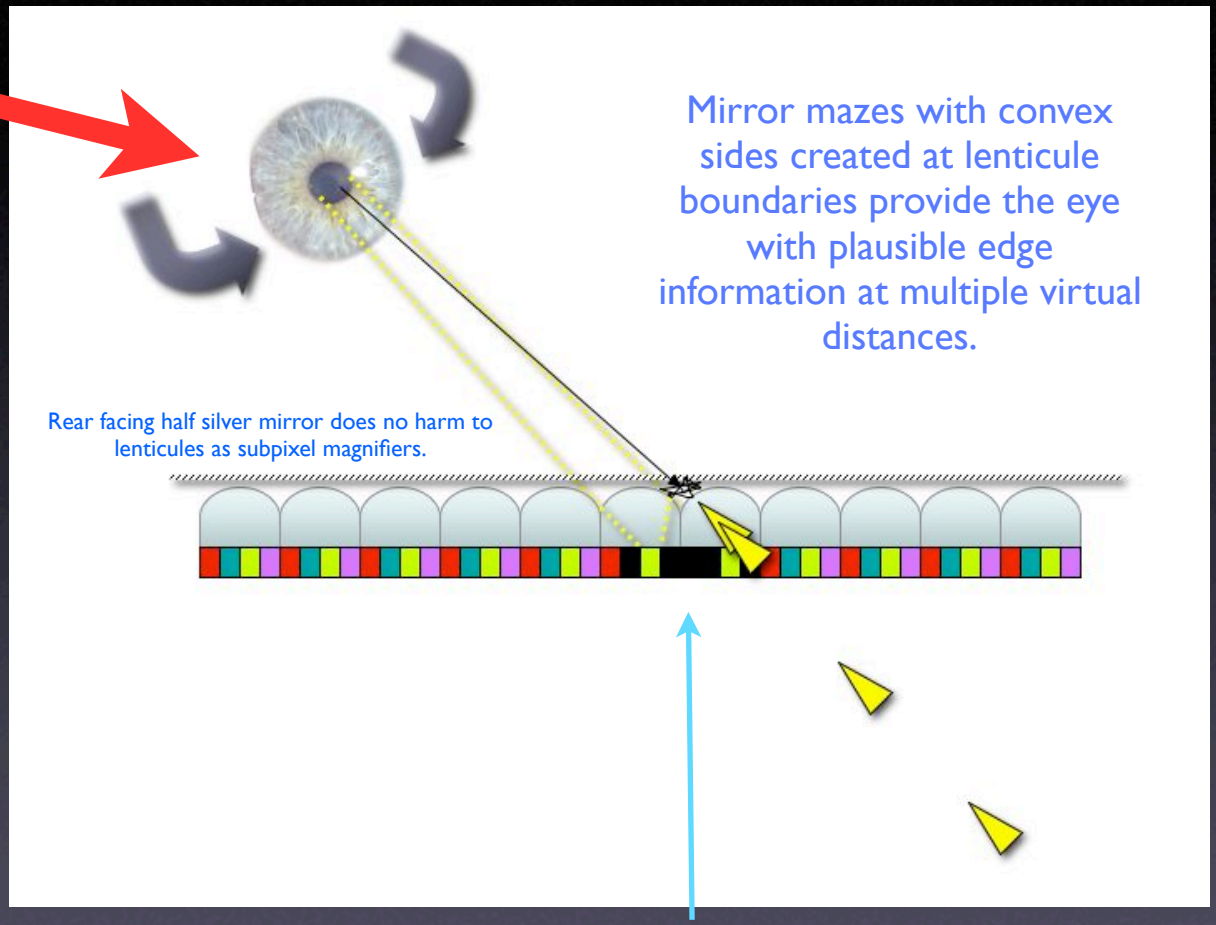


Note, however, that robust 3D Eye tracking and limited variation in head position relative to the display makes the use of fat lenticules with multi-sub pixels possible (ref Perlin,) and these might be turned into "Pancakelets."

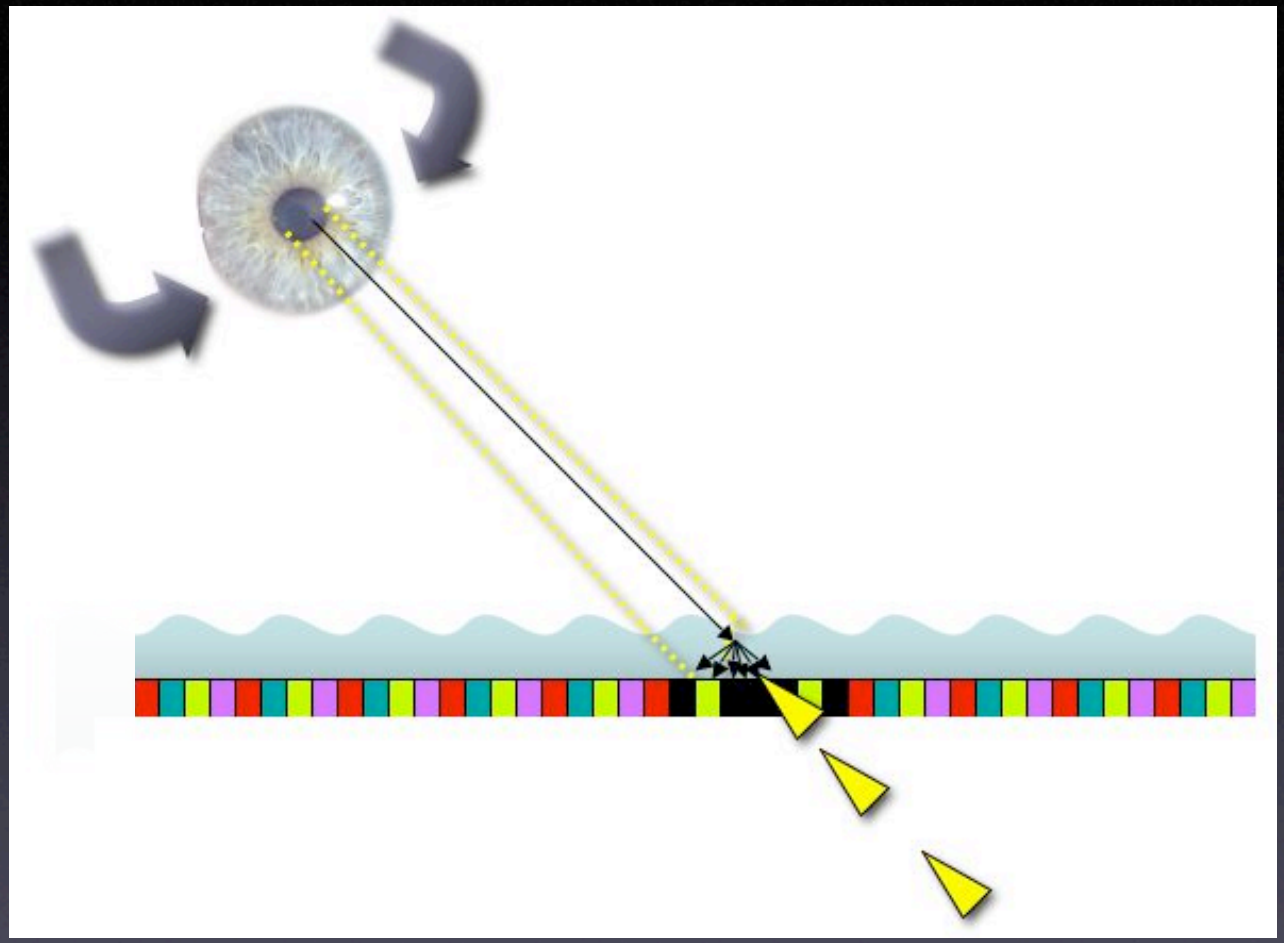
Another idea...

Warning! Even more speculative than other ideas in this presentation.

If this works (or not) it will teach us new things about human vision.



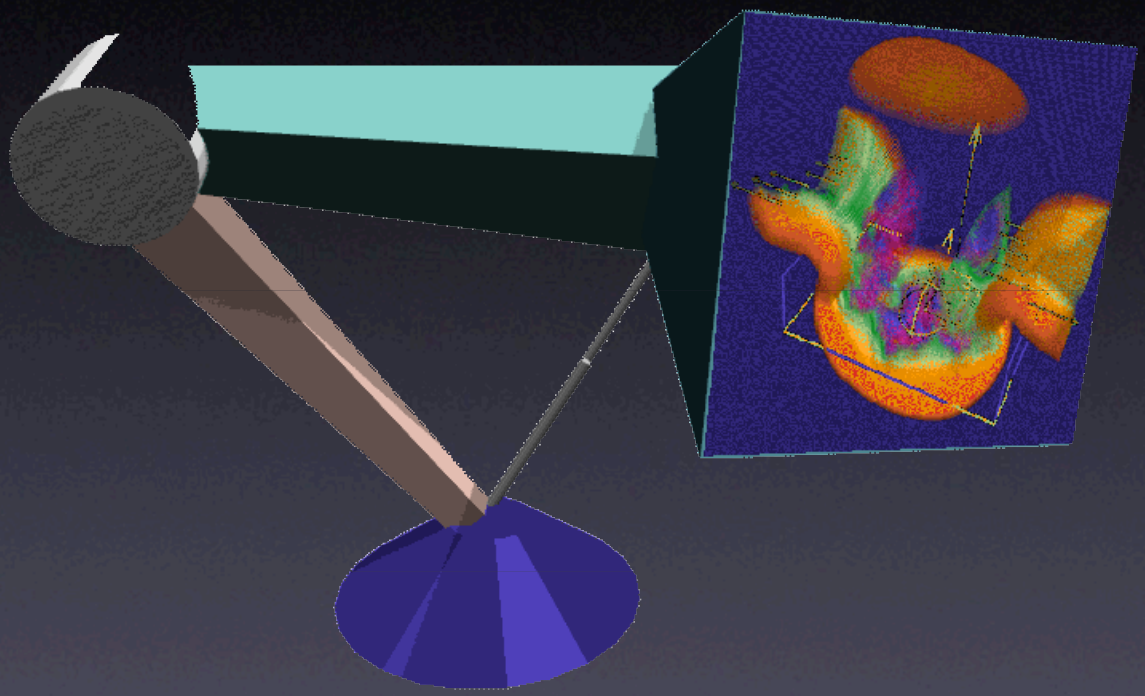
Pixels near depth cliffs are blacked out (so long as there is no eye in any of the the corresponding wedges) to avoid speckle-like artifacts.



Warning! More speculative than other ideas in this presentation.

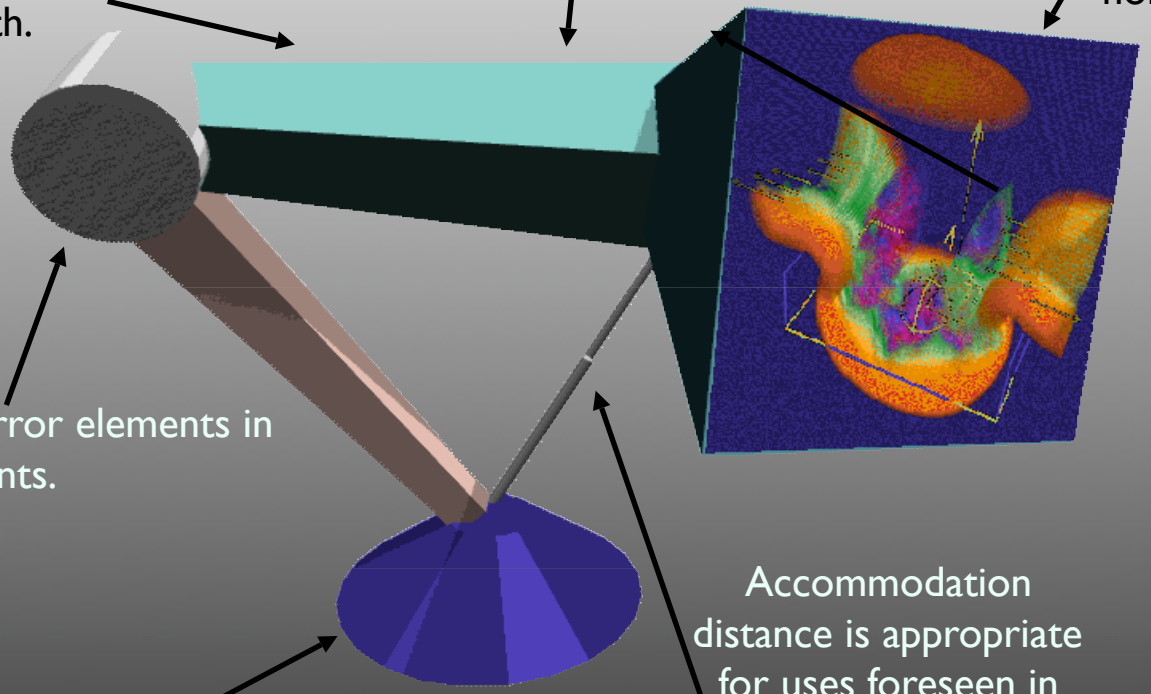
A related experiment would be to smooth the troughs between the lenticules to reduce the prominence of edges in the surface of the display.

Now let's consider the transmissive case.



Cododex suddenly looks completely different!
What's going on?

Perspective drawing of the transmissive variation of cocodex with parts labeled.



The cocodex arm is now hollow, because it contains the optical path.

Arrangement of three arm segments allows for a full and fast range of head tracking.

Screen surface is a transmissive holographic optical element.

Each eye gets its own exit pupil.

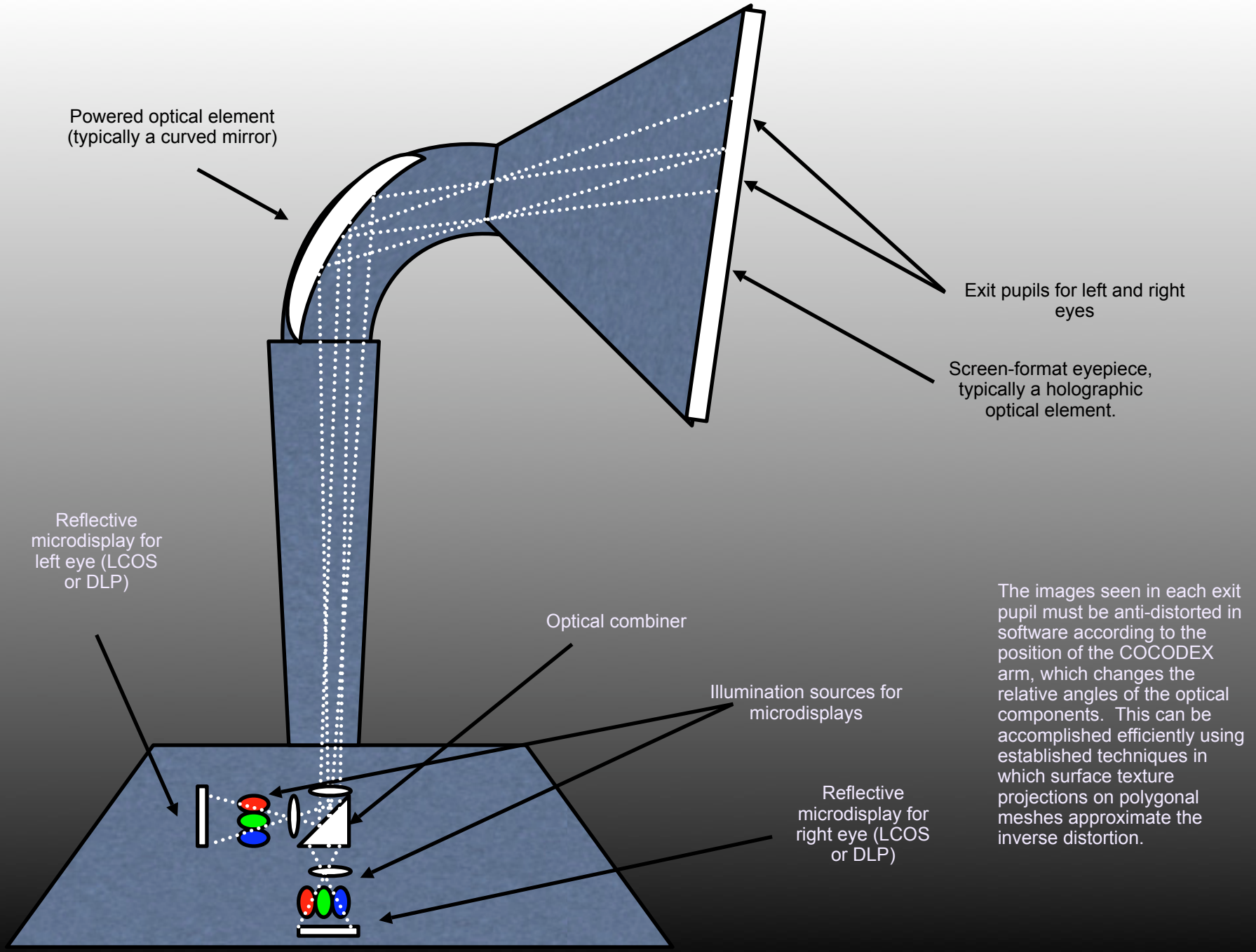
(Low) powered mirror elements in the joints.

Accommodation distance is appropriate for uses foreseen in this presentation.

Note that in this illustration, only the visual display is depicted, not other components such as cameras, speakers, light sources, and microphones.

The base contains two LCOS or DLP microdisplays (one for each eye,) illuminated by LEDs or lasers, and merged.

Telescoping redundant support arm for stability and to reinforce structure (which is weaker than the non-hollow variant.)



The images seen in each exit pupil must be anti-distorted in software according to the position of the COCODEX arm, which changes the relative angles of the optical components. This can be accomplished efficiently using established techniques in which surface texture projections on polygonal meshes approximate the inverse distortion.

Advantages of transmissive case:

- Full resolution of microdisplay parts are preserved.
- Potential accommodation problem is made moot.
- Probably would be bright, high contrast; beautiful.

Disadvantages of transmissive case:

- Only one user per cocodex.
- Bigger arm.
- Less potential use of off-the-shelf optical parts.

Note that there have been plenty of multiple exit pupil transmissive autostereo experiments but they were all stymied because they'd only work if the person's head barely moved; or else they would require big exit pupils, which meant impractical, giant optics. Cocodex in motion once again presents a potential way out of a long-standing dilemma.

Part Seven:
Summary of Risks and Potential

What are COCODEX's most significant vulnerabilities?

1) Will cocodex wallop you (or even someone else?)

2) Will Compound Portraiture meet human factors requirements?

3) Will it be too disruptive to have robotic moving objects in a work environment?

4) Will robotics be fast enough?

5) Will cocodex lose track of you too often?

6) Will pseudo-immersion work?

7) Will accommodation be a problem?

What are COCODEX's most significant strengths?

1) Solves full duplex tele-immersion problem.

2) Desktop design; No need for special room.

3) Supports >2 users.

4) Supports normal range of human motion while seated.

5) Uniquely reduces requirements so that known parts can already perform well enough to meet known human factors specs.

6) Supports heterogeneous and evolving information technology practices through idea of "Pseudo-immersion." (You can use a physical pc, phone, etc. along with virtual hi res display walls and command centers at the same time as you use cocodex as a tele-immersion device.)

7) For above reasons has potential for rapid and widespread adoption, unlike known alternatives.

8) Provides improved UI for working with volumetric data.

9) Supports reconstruction strategies that can benefit from predictive filtering to reduce apparent latency during long distance collaborations.

So that's what COCODEX is.