electronic IMAGING JANUARY 2023

Display system sharpness modeling and requirement in VR and AR applications

Jiawei Lu[1], Trisha Lian[1], Jerry Jia[1]

[1]Meta Reality Labs (California & Washington)

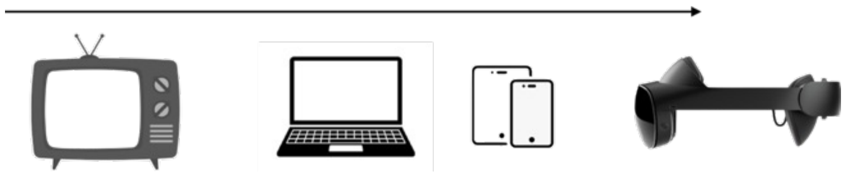
Meta

Context

… you've got to start with the customer experience and work backwards for the technology."

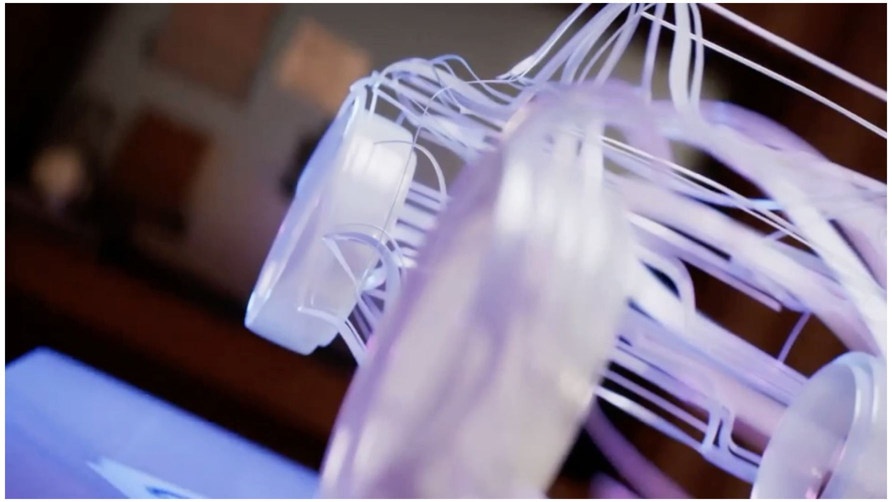
Steve Jobs

…It's increasingly so when devices are getting closer to human senses

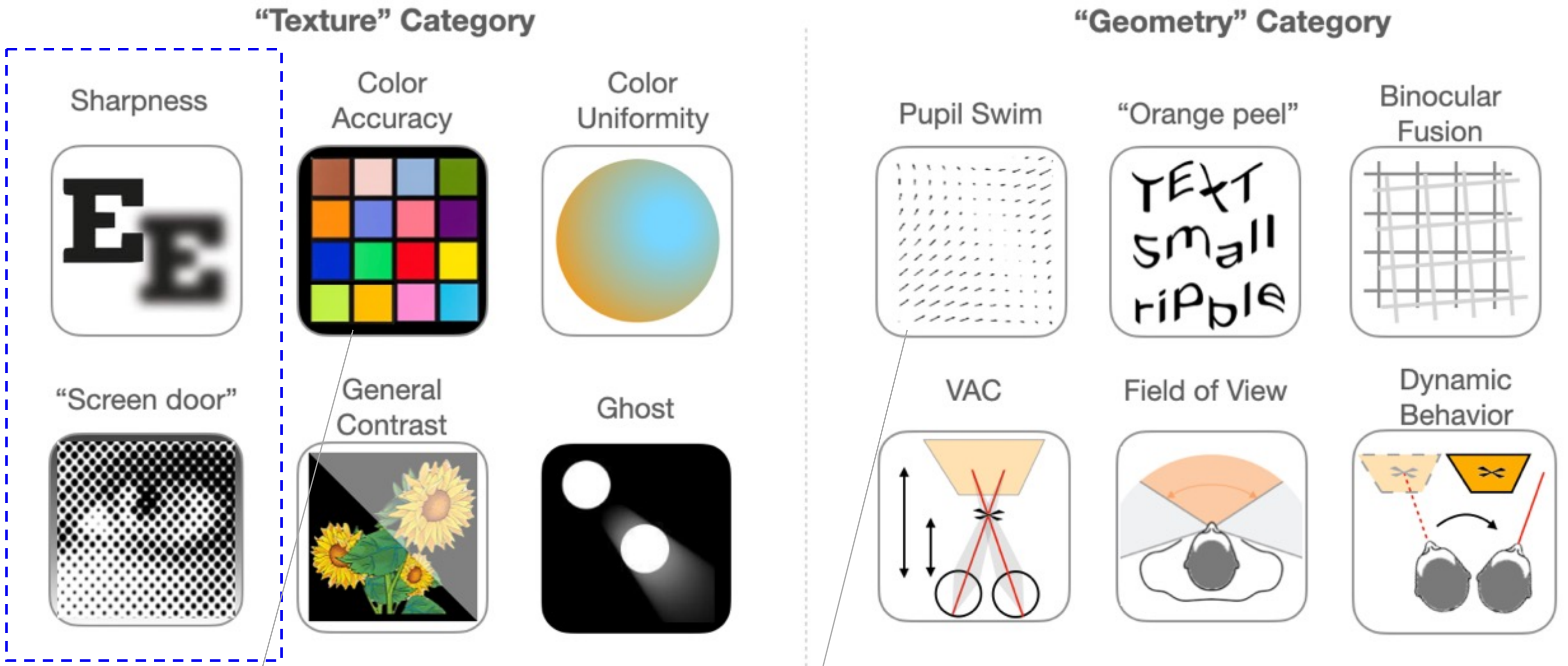


Meta newest headset: Quest Pro

MR is a feature: "Collaborate in mixed reality together, apart"



What experiences matter? (Product is a balance of all)



Today's topic

...and more

M. Wei, W. Bao, Z. Huang, J. Oberländer, S Rüffer, and J. Jia, "Adaptive Display White Point for Enhancing Viewing Experience of Mixed Reality Headsets" Color Imaging Conference 2022

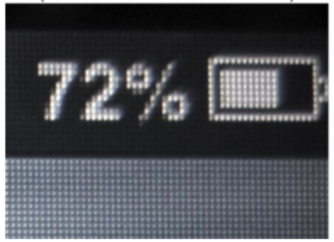
T. T. Chan, Y. Wang, R. H. Y. So and J. Jia, "Predicting Subjective Discomfort Associated with Lens Distortion in VR Headsets During Vestibulo-Ocular Response to VR Scenes," in *IEEE TVCG* (Also invited talk at ISMAR 2022, invited talk at IDS 2022)

New Resolution Challenge for VR/MR/AR (display & camera)

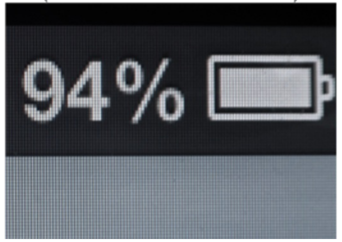
"Why do we care about resolution in post- Retina Display era?"

- The optics in XR effectively stretches display to cover a much larger field of view (e.g., iPhone 13: 28 deg wide (+/-14); VR challenge: >90 deg)
- Pixel per degree (PPD) is a key factor (not the only key factor): large FOV -> low PPD
- **So for XR: we need a lot more pixels, and a lot smaller pixels than a phone display!**
  - VR challenge: display approximately 10x more pixels on 1/4 size of iPhone display
- More layers are more challenging: MR is combining display and camera

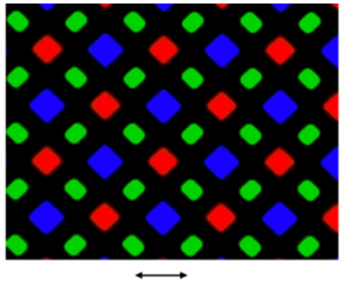
iPhone 3GS: 163ppi
(28PPD at 25cm distance)



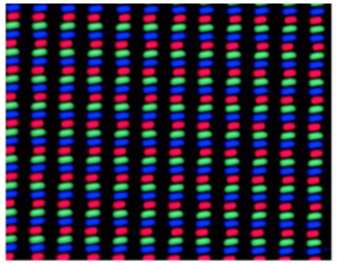
iPhone 4: 326ppi
(56PPD at 25cm distance)



iPhone 13: 460ppi
(80PPD at 25cm distance)

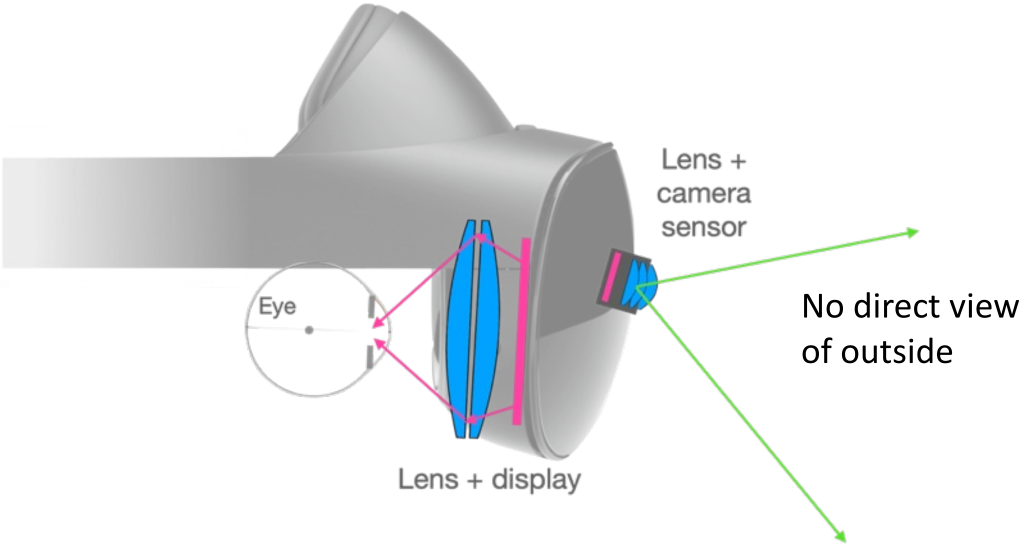


55um
iPhone 13

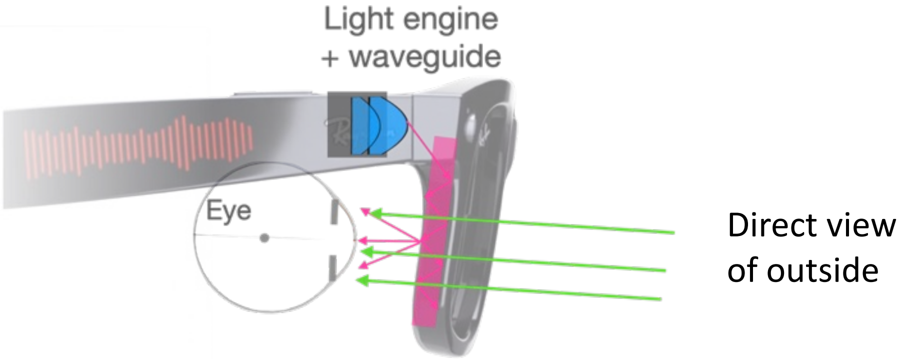


24um
VR today

VR/MR:

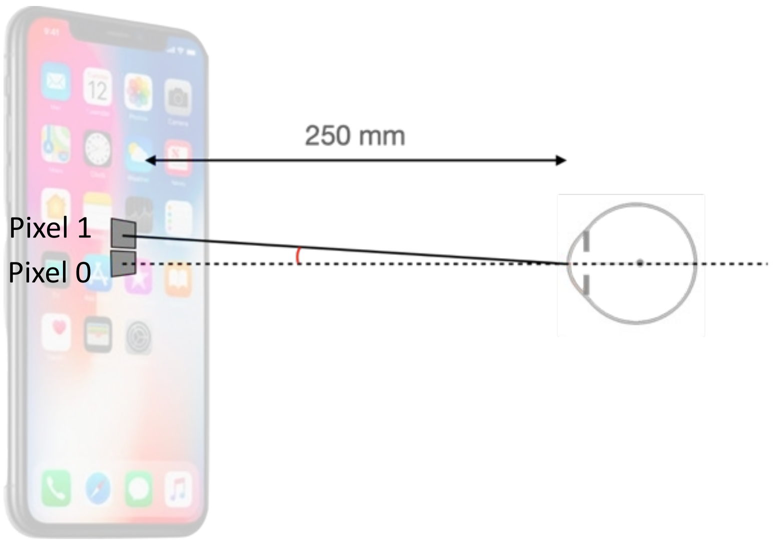


AR:



Eye operates in angle space (PPD matters, not PPI)

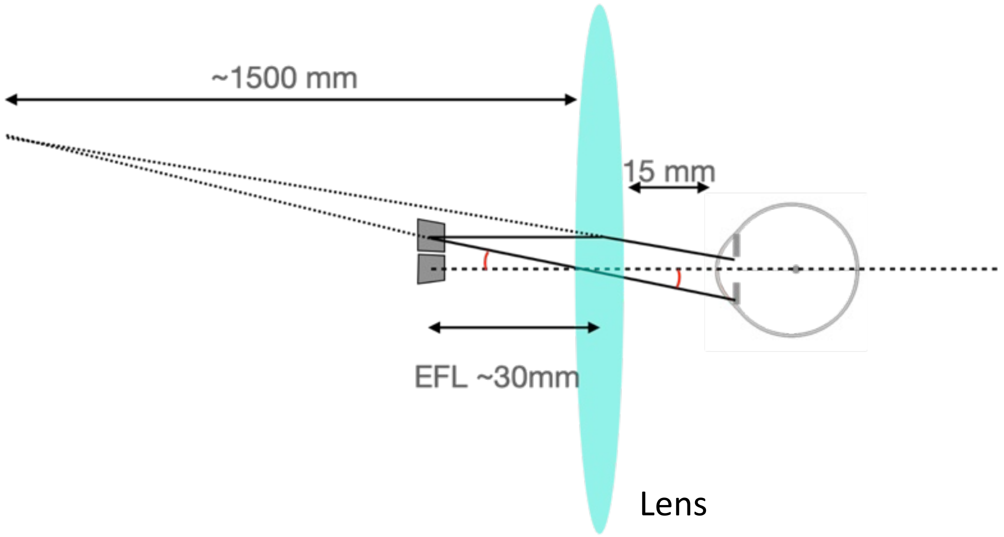
Direct view of display:



$$\text{Deg per Pixel} = atan\left(\frac{\text{50 um (1 pixel)}}{\text{250 mm (view distance)}}\right)$$

$$\text{Pixel per Deg (PPD)} = \text{Deg per Pixel}^{-1} = \text{87 PPD}$$

VR:



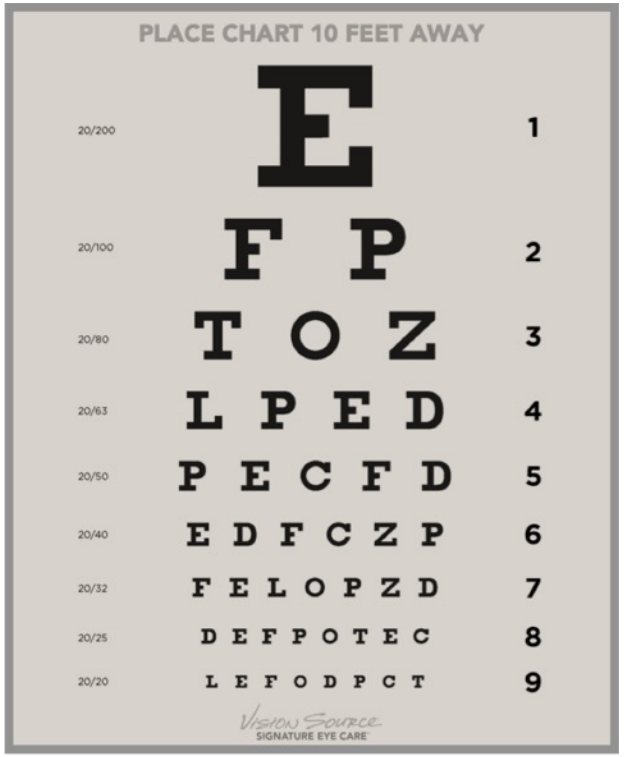
$$\text{Deg per Pixel} = atan\left(\frac{\text{50 um (1 pixel)}}{\text{30 mm (EFL)}}\right)$$

$$\text{Pixel per Deg (PPD)} = \text{Deg per Pixel}^{-1} = \text{10 PPD}$$

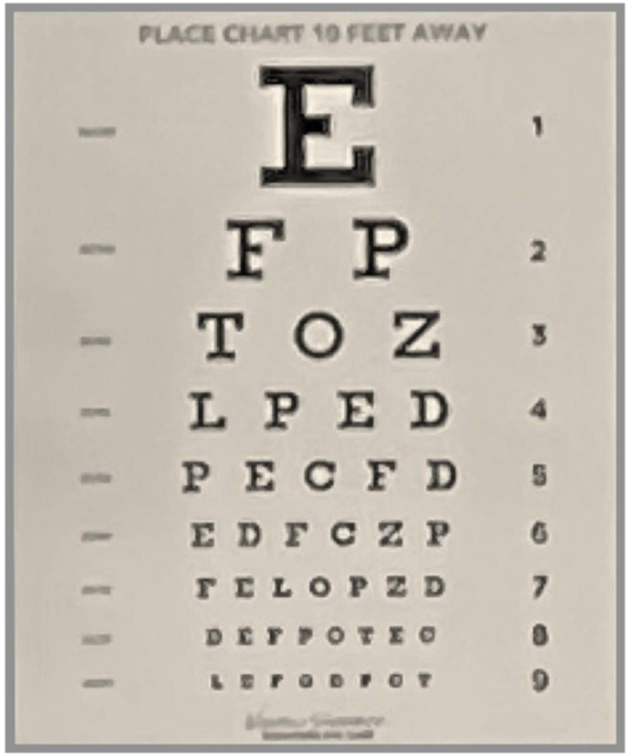
We need a lot denser pixels in VR!

What’s a good PPD – it depends

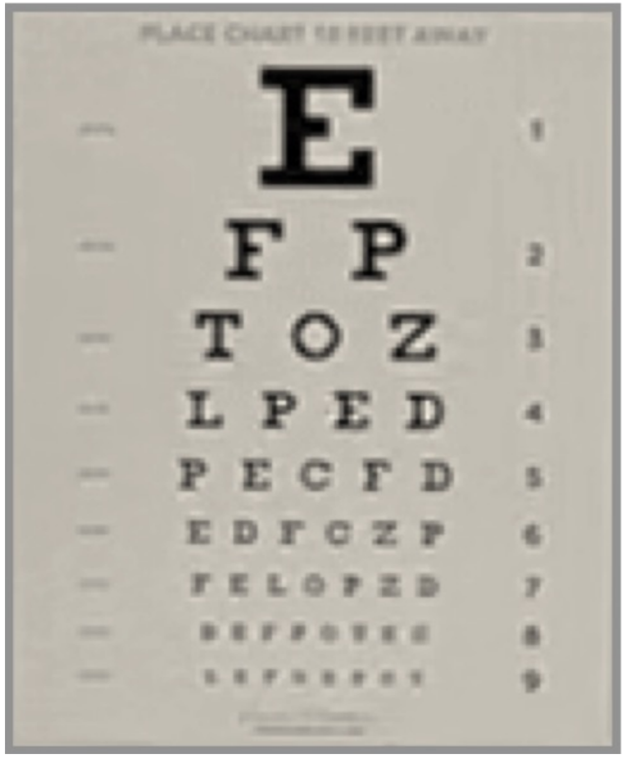
- Do we need 60PPD for “20/20 vision”? (yes)
- Do we need 60PPD for all applications? (no)



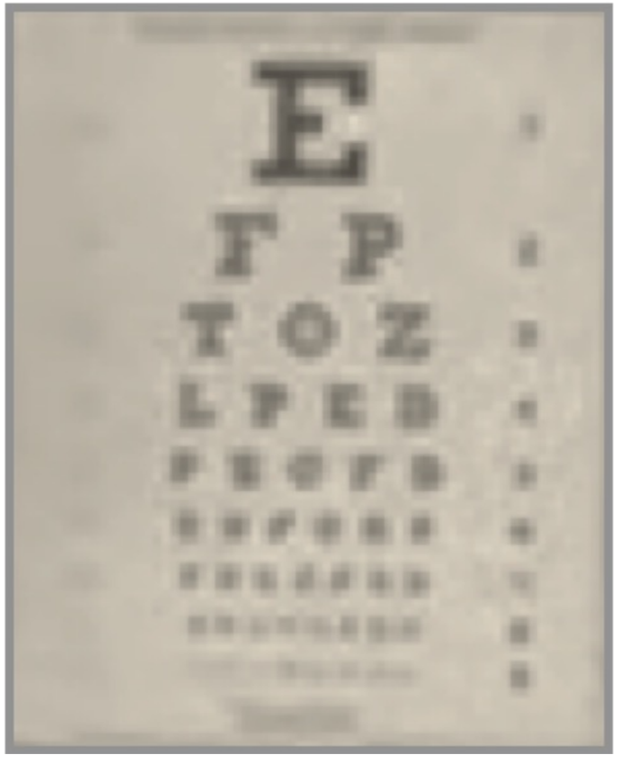
Reference photo
(>> 20/20)



iPhone 13 photo
56PPD. (~20/25)

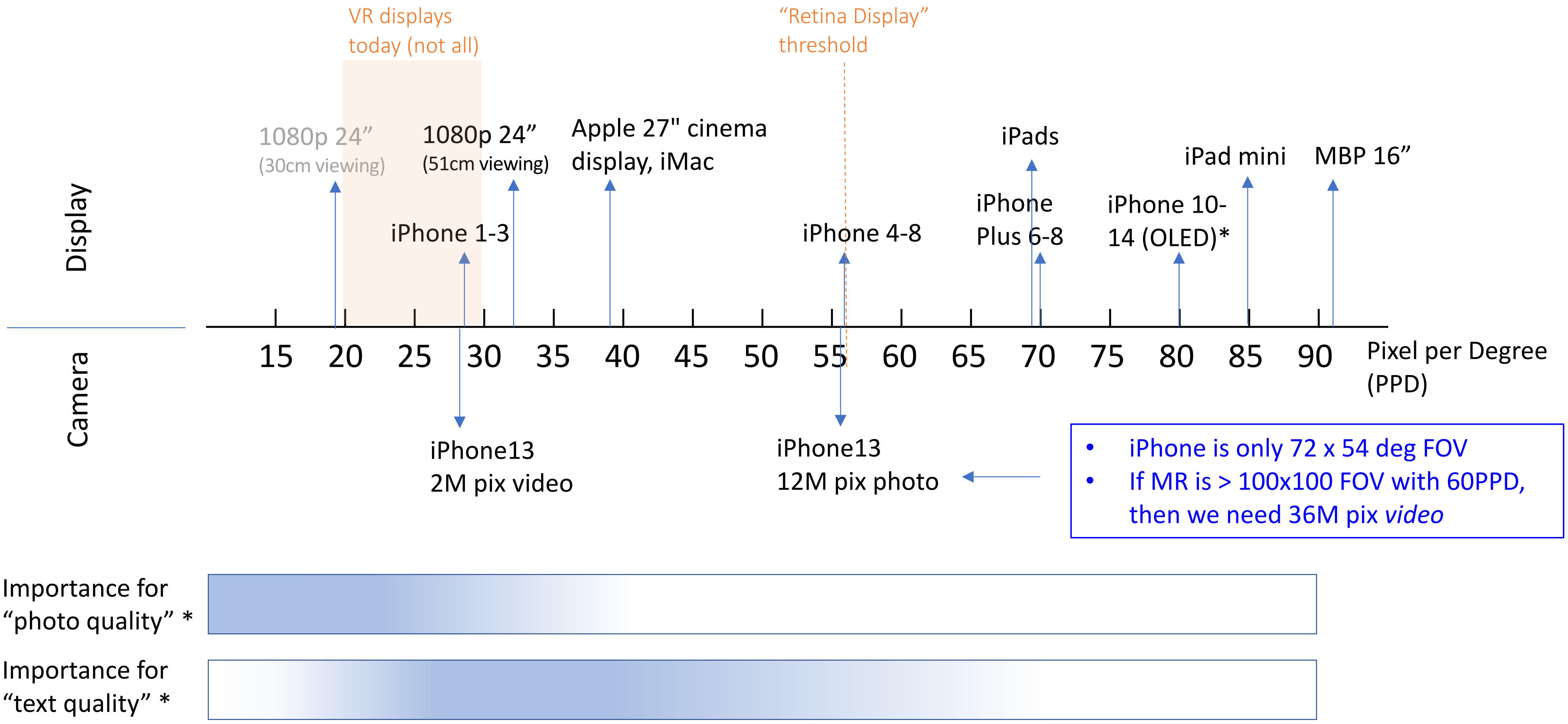


iPhone 13 video frame
28PPD. (~20/40)



iPhone 13 video frame
15PPD. (~20/80)

The PPD landscape and "engineering reality"



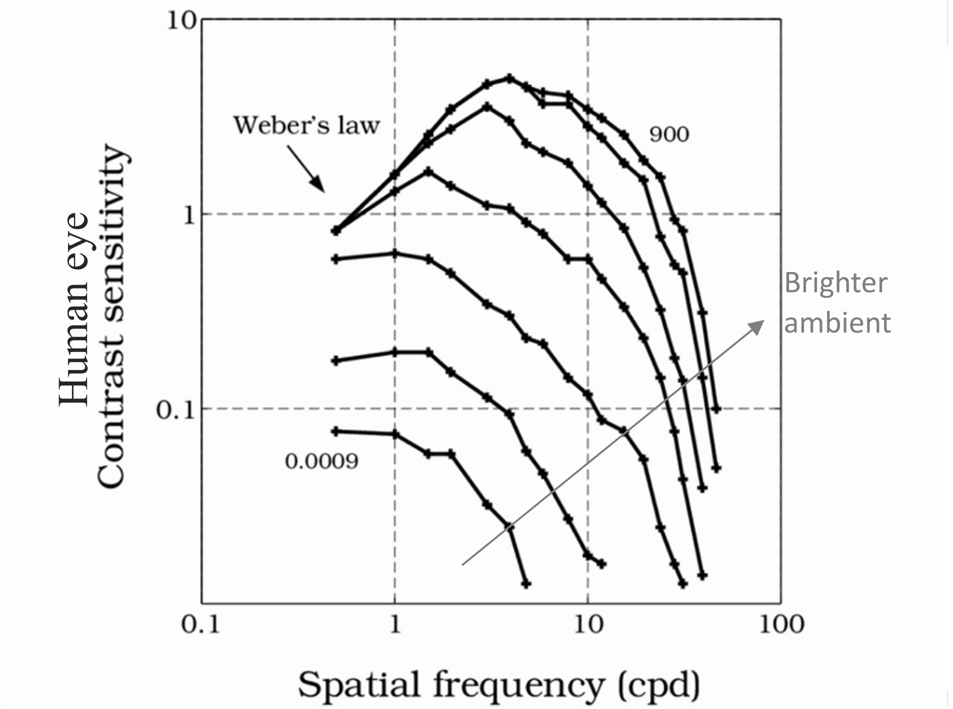
- iPhone is only 72 x 54 deg FOV
- If MR is > 100x100 FOV with 60PPD, then we need 36M pix *video*

* Design heuristics

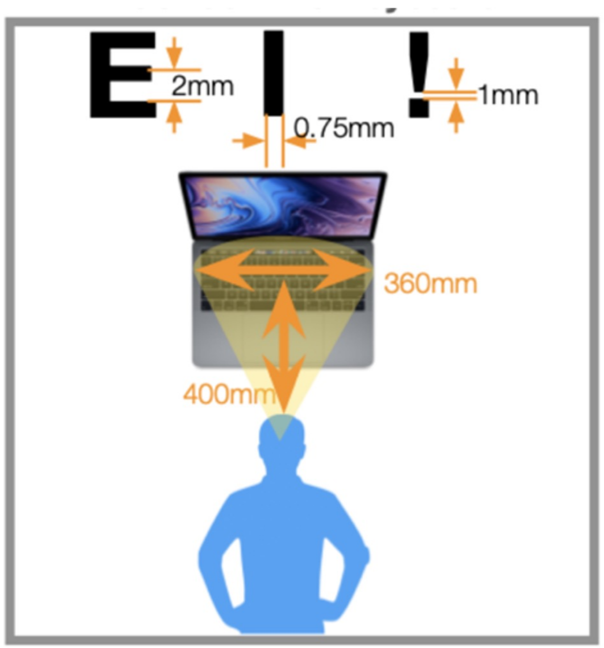
Content-driven vs. human contrast sensitivity function

- CSF is a reference; don’t overuse it: not recommend to integrate w/ frequency
- Important: 1) content is king (photograph vs. text); 2) “legibility” vs. “quality”

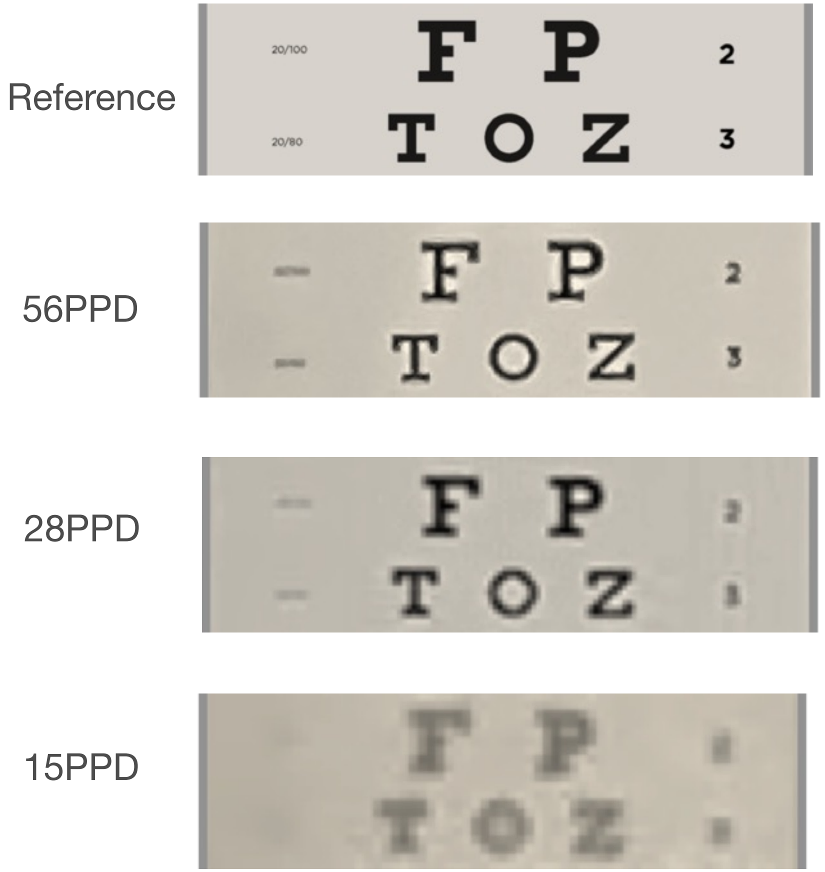
CSV is a reference. Content is more important



Example: Keyboard *“legibility”* (~ 2nd row of eye chart) needs spatial frequency of ~5 cycles/deg (10PPD)



“Text *quality*” requires >> 10 cy/deg



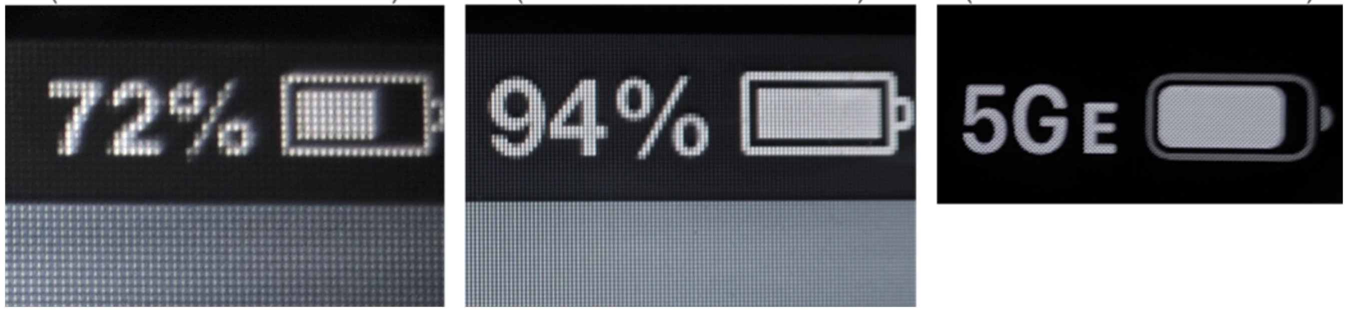
One more thing: it’s not *only* about “resolution” or PPD

1. PPI or PPD (resolution)

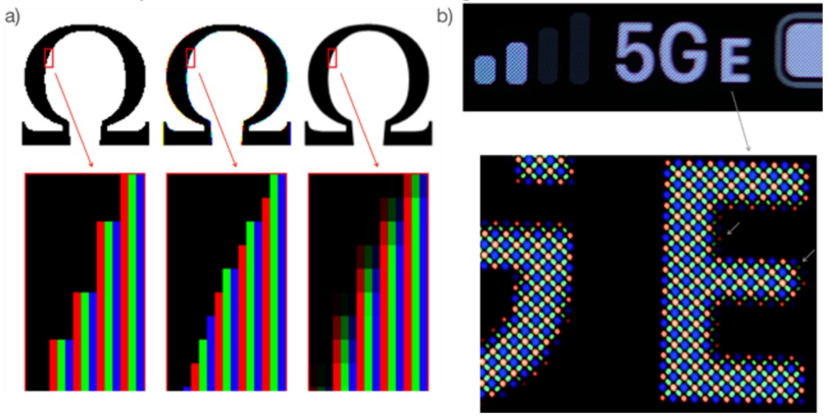
iPhone 3GS: 163ppi (28PPD at 25cm distance)

iPhone 4: 326ppi (56PPD at 25cm distance)

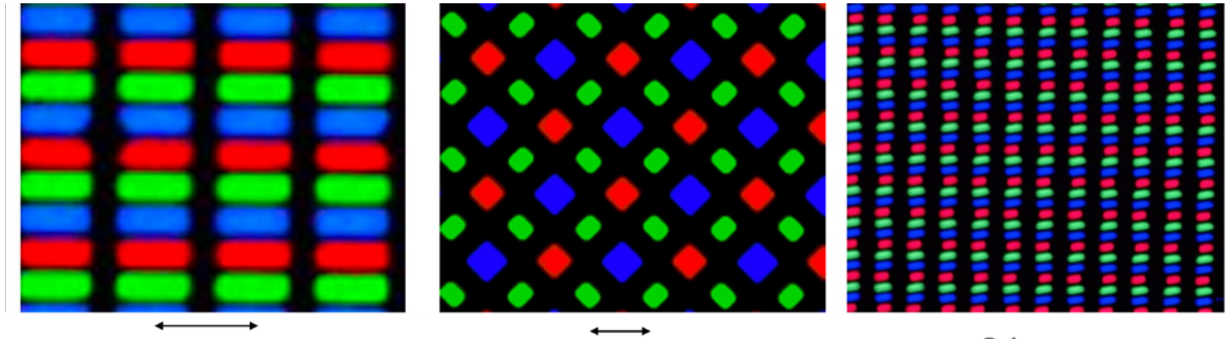
iPhone 13: 460ppi (80PPD at 25cm distance)



3. Graphics/rendering



2. Pixel shape, fill factor

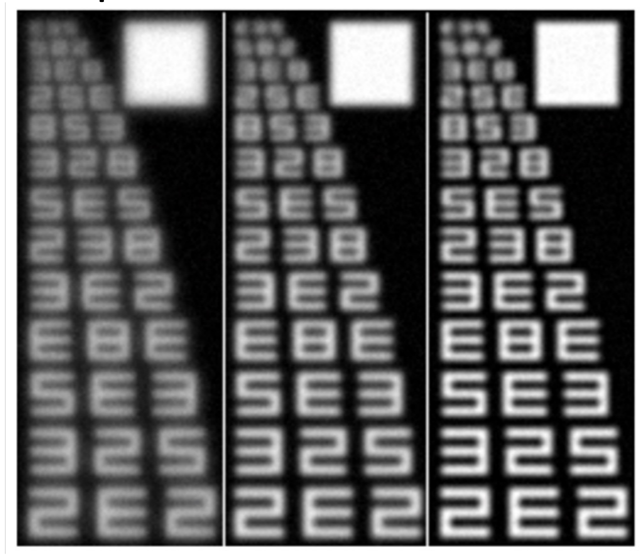


78um
iPhone 4

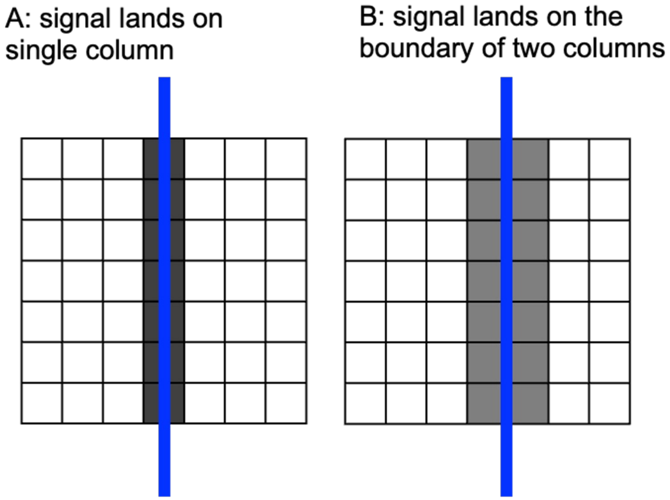
55um
iPhone 13

24um
Arcata

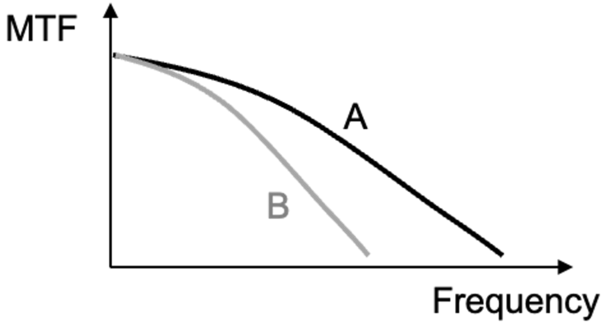
4. Optical blur



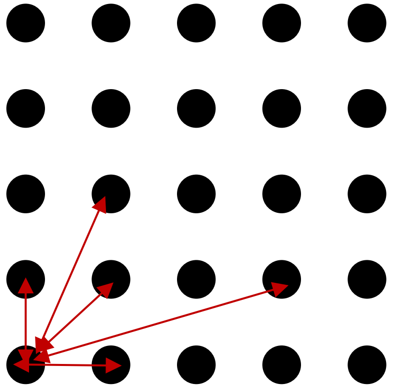
Further More: nature of sampling

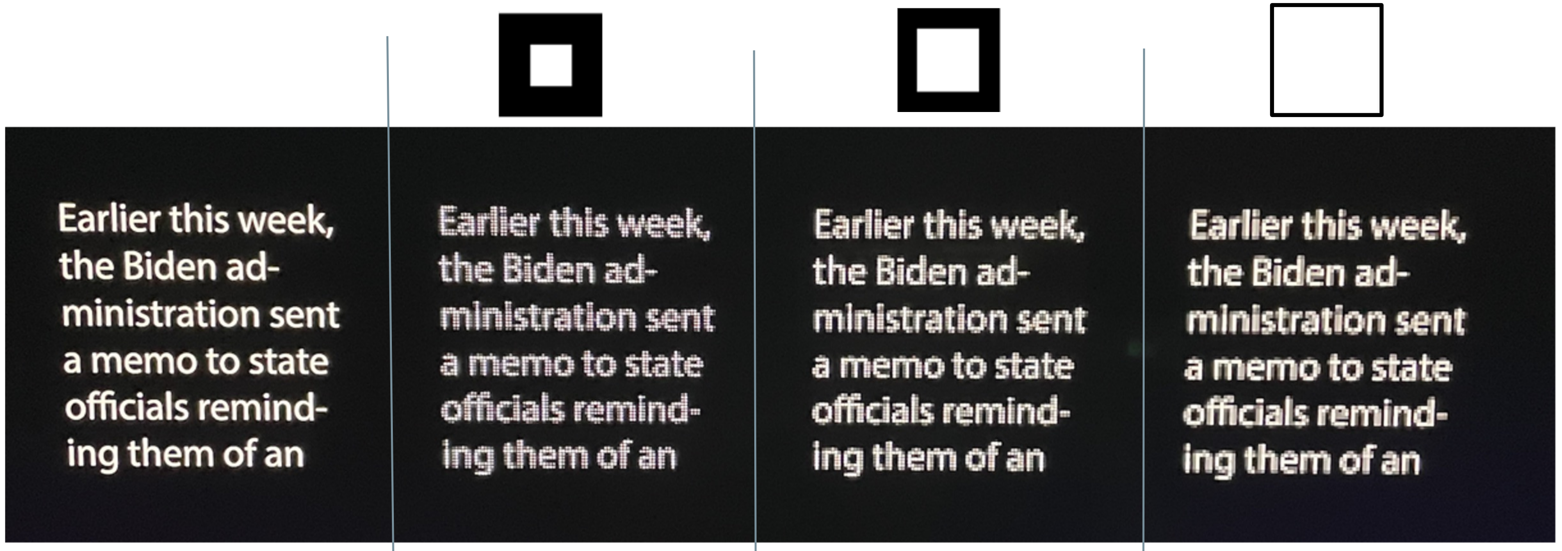


1. Sampling location



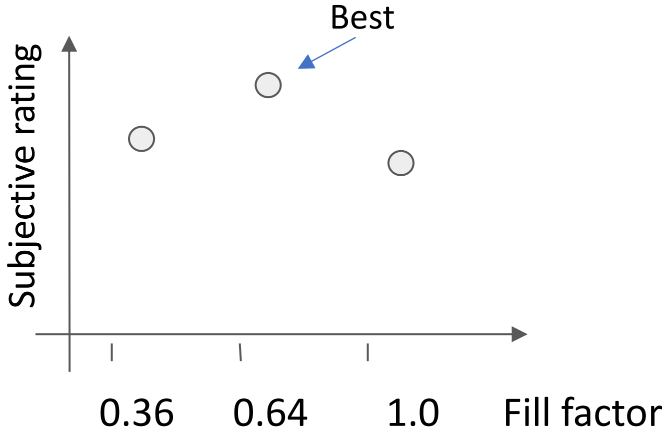
2. Sampling direction





Original content | 36% fill factor | 64% fill factor | 100% fill factor

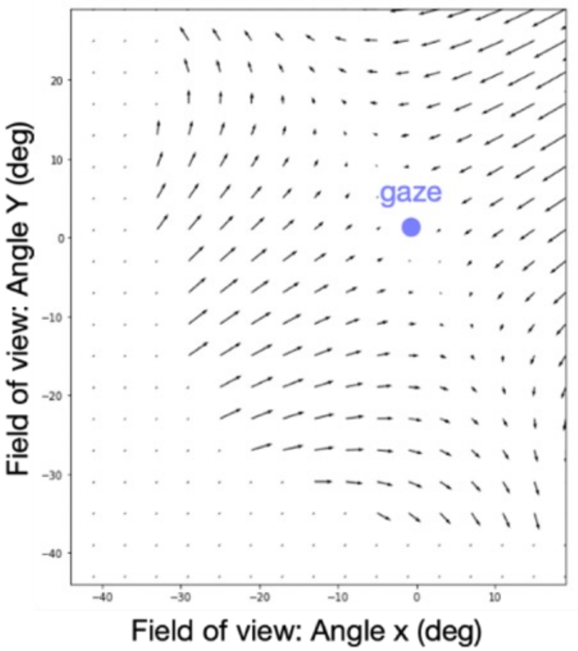
3. Pixel fill factor



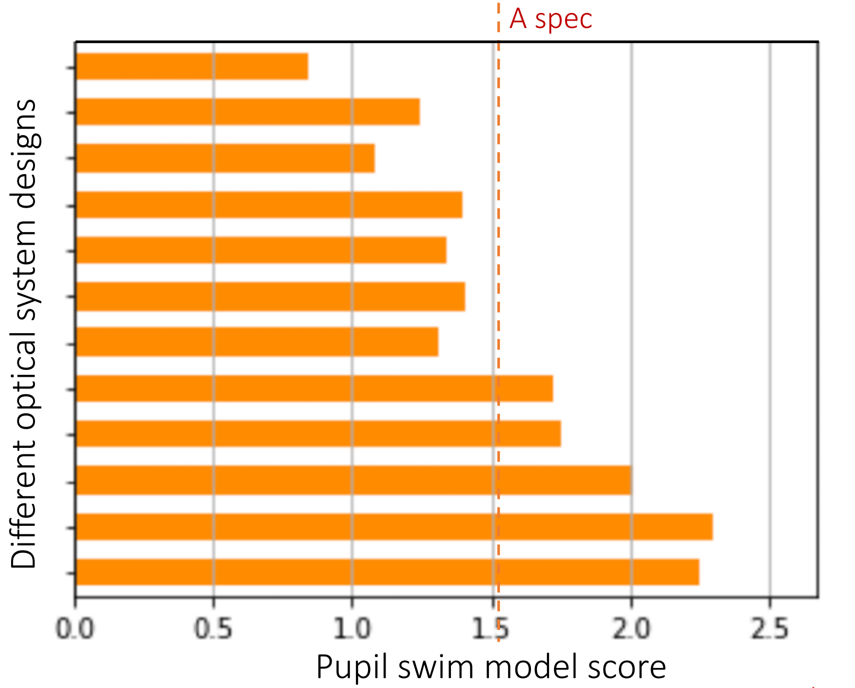
Lets take a pause: What are we looking for?

- Engineering challenges: both display and camera need a lot *more* pixels, a lot *bigger* FOV
- User experience/content need: 1) photographic, 2) text quality
- Design challenges: It's not just PPD, there are *many* variables such as fill factor
- What we are looking for: A model to quantify all factors above, and a spec

An example of mathematical model of "motion sickness" for VR:



Incredibly helpful in design and architecture explorations in product development



Jerry Jia *IEEE TVCG*

Give MTF a chance

*Question: Is there a way to **capture** all the complex contributors, **match** user experience, and **guide** practical product design?*

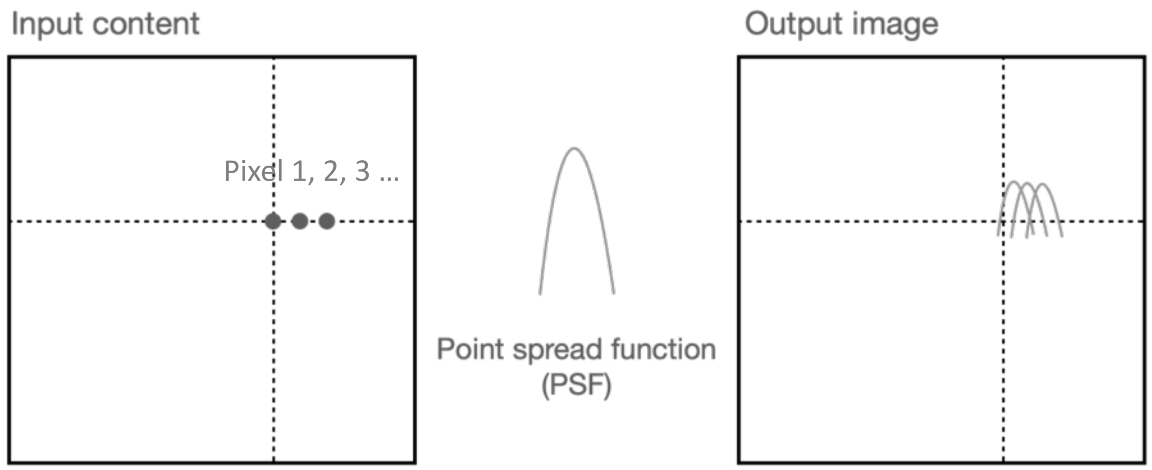


Image formation (convolution in spatial domain):

- Every pixel in input content results in a shifted PSF. The output image is the overlapped sum of these PSFs.
- Each point in output image receives contribution from many pixels in input content
- (Translate to frequency domain): PSF -> line spread function (LSF) -> MTF

- MTF is a popular metric in frequency domain to measure image/display system
- Conveniently Multipliable: MR System MTF = Camera MTF x VR MTF = (Lens MTF x Sensor MTF) x (Lens MTF x Display MTF)
- Quantifiable with spec: “at about 3-15 cycles/deg, we require > 20% ideally > 35% contrast”
  - Traditional wisdom: for visual usage, when contrast drops below 20% - image looks blurry; When “image looks good” the MTF is > 35% @useful frequency

Our solution: "system MTF" model and metric

*Question: Is there a way to **capture** all the complex contributors, **match** user experience, and **guide** practical product design?*

$$SystemMTF = \int_0^{40}\int_0^{\pi}\frac{1}{p}\int_{-p/2}^{p/2}\left[\left(H_{gh}(\xi,\theta)e^{-j2\pi\xi t}\right) * \sum_n \delta\left(\xi - n\frac{1}{p(\theta)}\right) \times PIX(\xi,\theta)\right]H_{opt}(\xi)w(\theta)w(\xi)\,dt\,d\theta\,d\xi$$

Graphics/ rendering

"Shift invariant"

PPD/ resolution

Pixel shape/ fill factor

Optics & eye blur

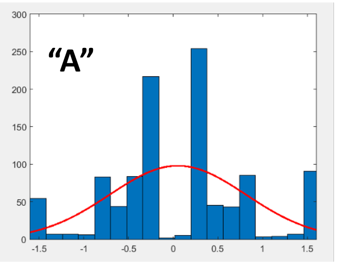
Weighting over orientation & frequency

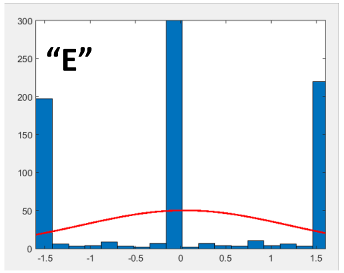
- The best way to achieve all above 3 goals is through math modeling
- We did not start from nothing: we expanded traditional MTF theory (point source + Fourier transform) with special mathematical treatments:
  - How to make the system MTF shift-invariant
  - Weighted over content frequency and orientation for text quality

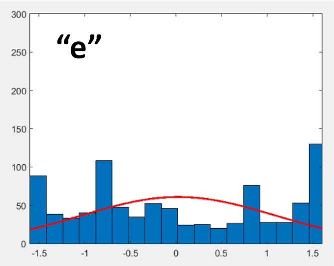
Orientation: English text is dominantly vertical

Aa Bb Cc Dd Ee
Ff Gg Hh Ii Jj Kk
Ll Mm Nn Oo Pp
Qq Rr Ss Tt Uu
Vv Ww Xx Yy Zz

Orientation analysis





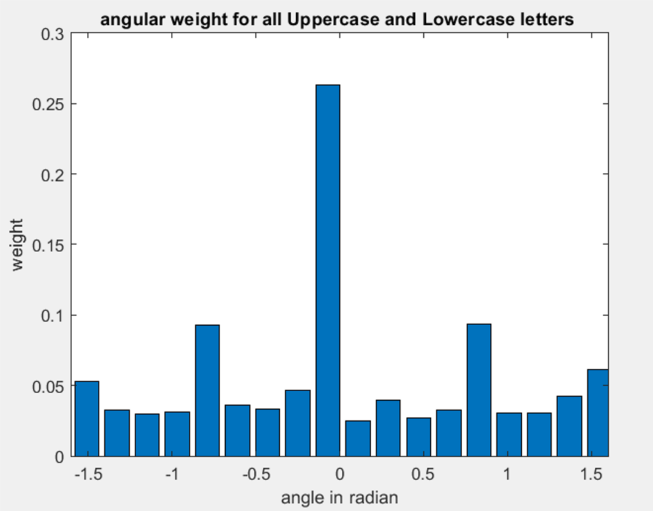


Average over

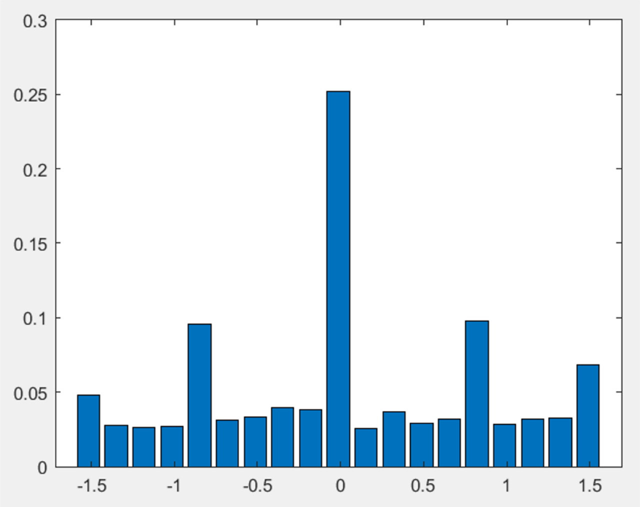
Letters' usage frequency

Table 1
Raw Case-Sensitive Single-Letter Counts from the NYT Corpus

Letter	Uppercase *f*	Lowercase *f*	Uppercase Rank	Lowercase Rank
A	280,937	5,263,779	3	3
B	169,474	866,156	8	20
C	229,363	1,960,412	5	12
D	129,632	2,369,820	12	11
E	138,443	7,741,842	11	1
F	100,751	1,296,925	17	15
G	93,212	1,206,747	19	17
H	123,632	2,955,858	13	9
I	223,312	4,527,332	6	6
J	78,706	65,856	20	25
K	46,580	460,788	22	22
L	106,984	2,553,152	15	10
M	259,474	1,467,376	4	14
N	205,409	4,535,545	7	5
O	105,700	4,729,266	16	4
P	144,239	1,255,579	10	16
Q	11,659	54,221	24	26
R	146,448	4,137,949	9	8
S	304,971	4,186,210	2	7
T	325,462	5,507,692	1	2
U	57,488	1,613,323	21	13
V	31,053	653,370	23	21
W	107,195	1,015,656	14	19
X	7,578	123,577	25	23
Y	94,297	1,062,040	18	18
Z	5,610	66,423	26	24

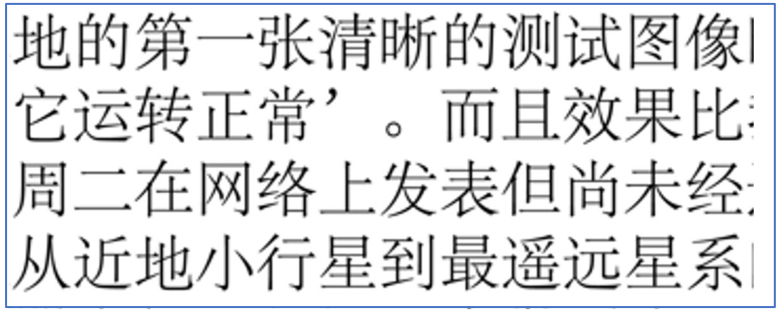


Average over Fonts/ sizes

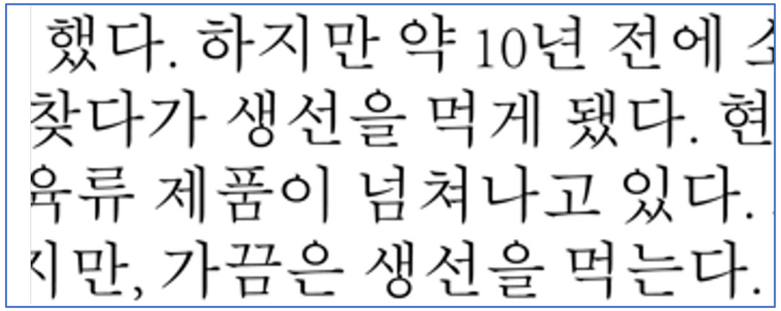


Different languages have different orientations

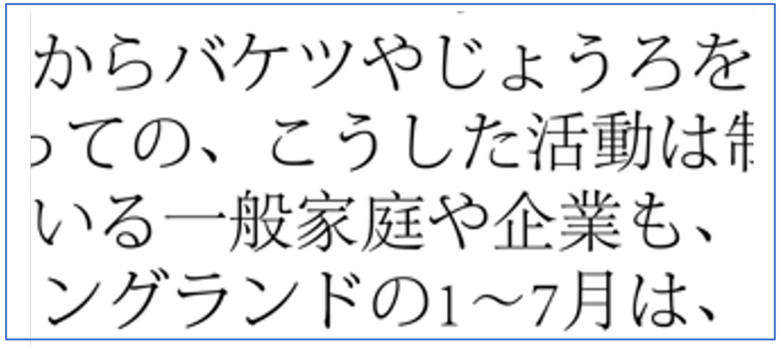
Chinese

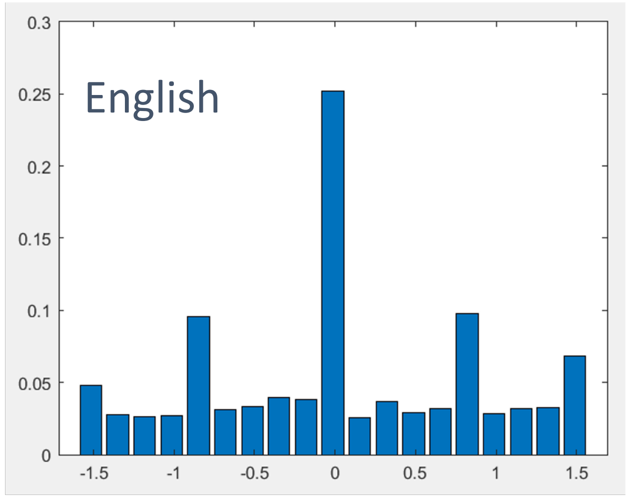


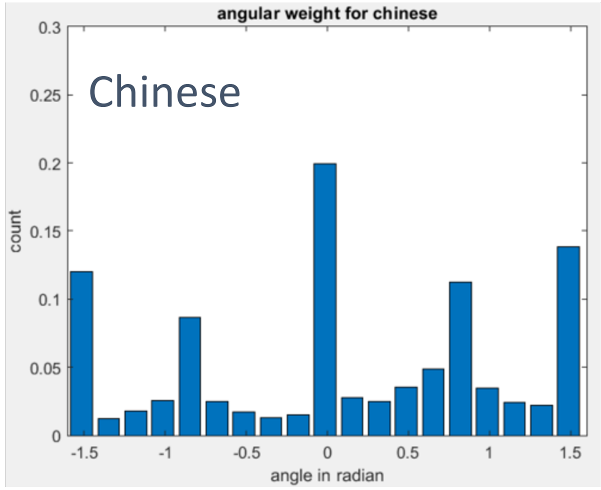
Korean

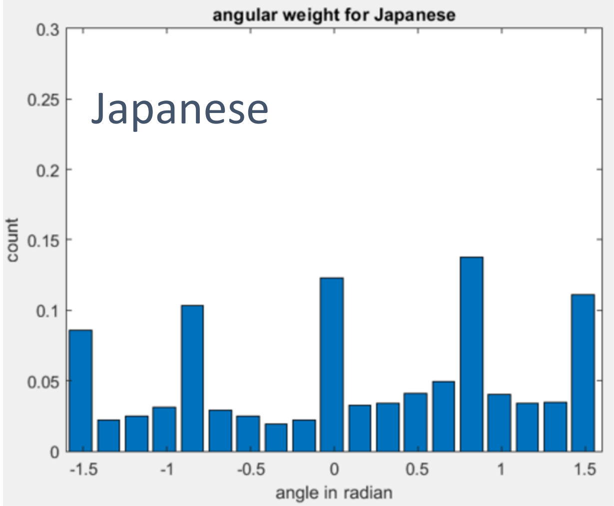


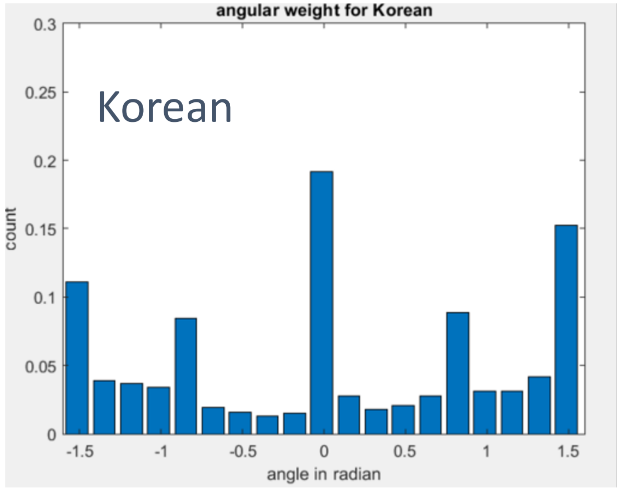
Japanese









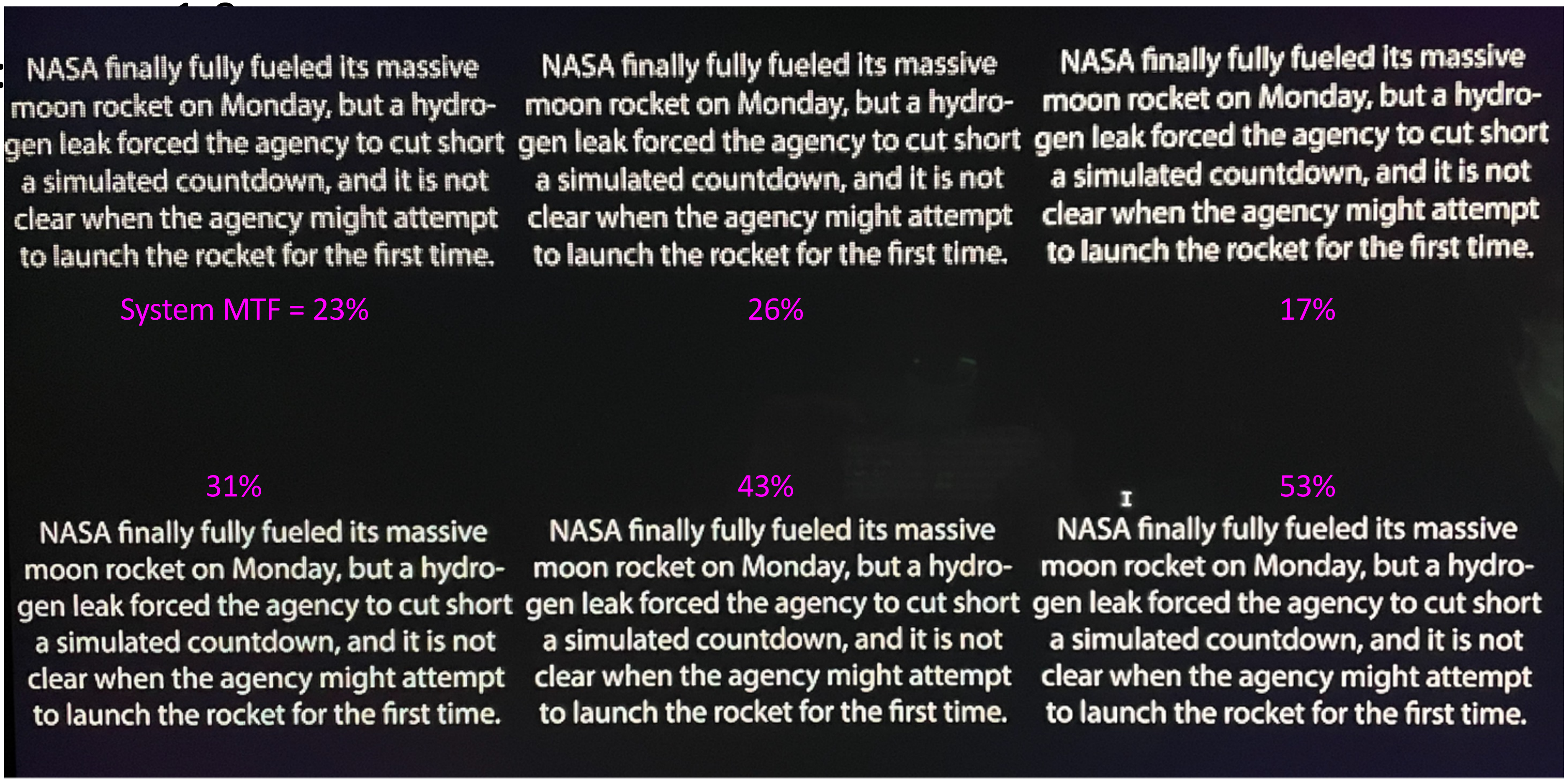


Demo : Fill factor is a strong contributor in addition to PPD

FF: 0.36 0.64

PPD:

25

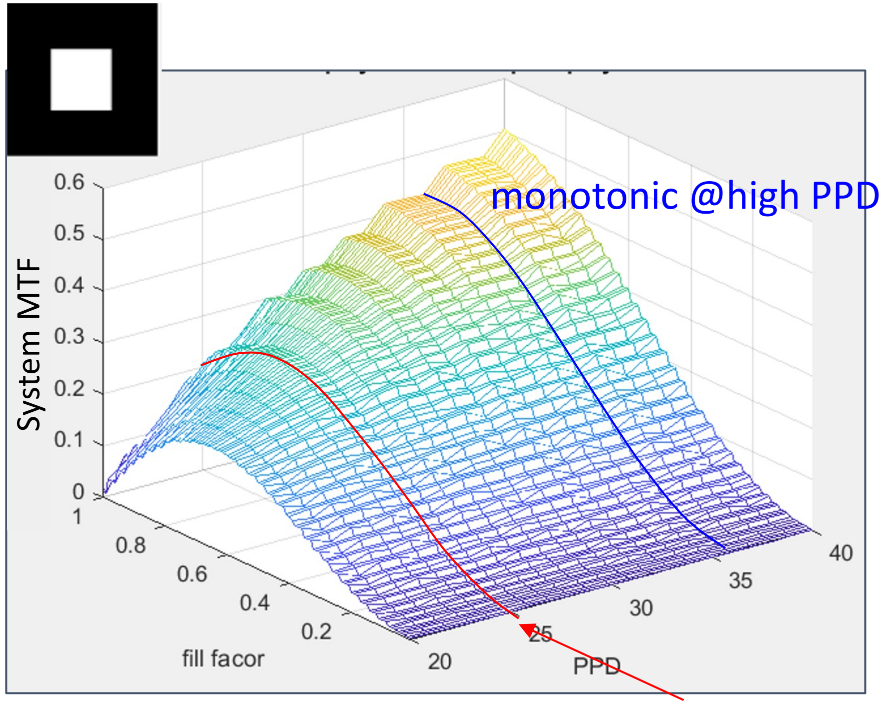


36

Photo capture of emulation on a high-res display

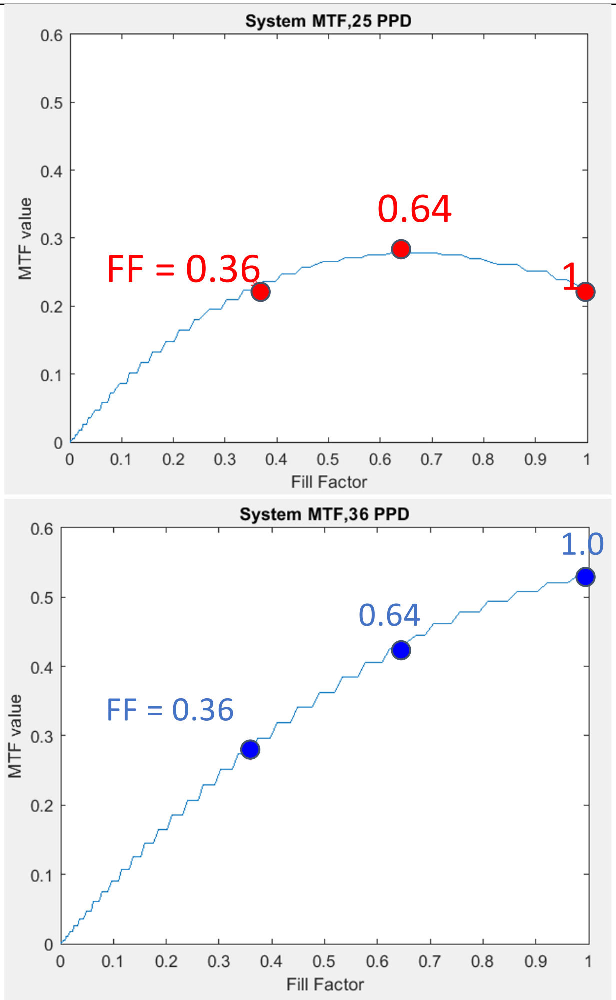
Display experience depends on PPD *and* fill factor

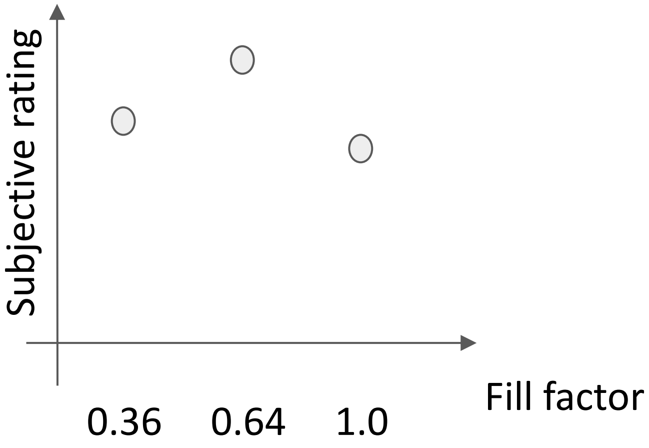
Test Frequency: 15.5 cy/deg



25 PPD

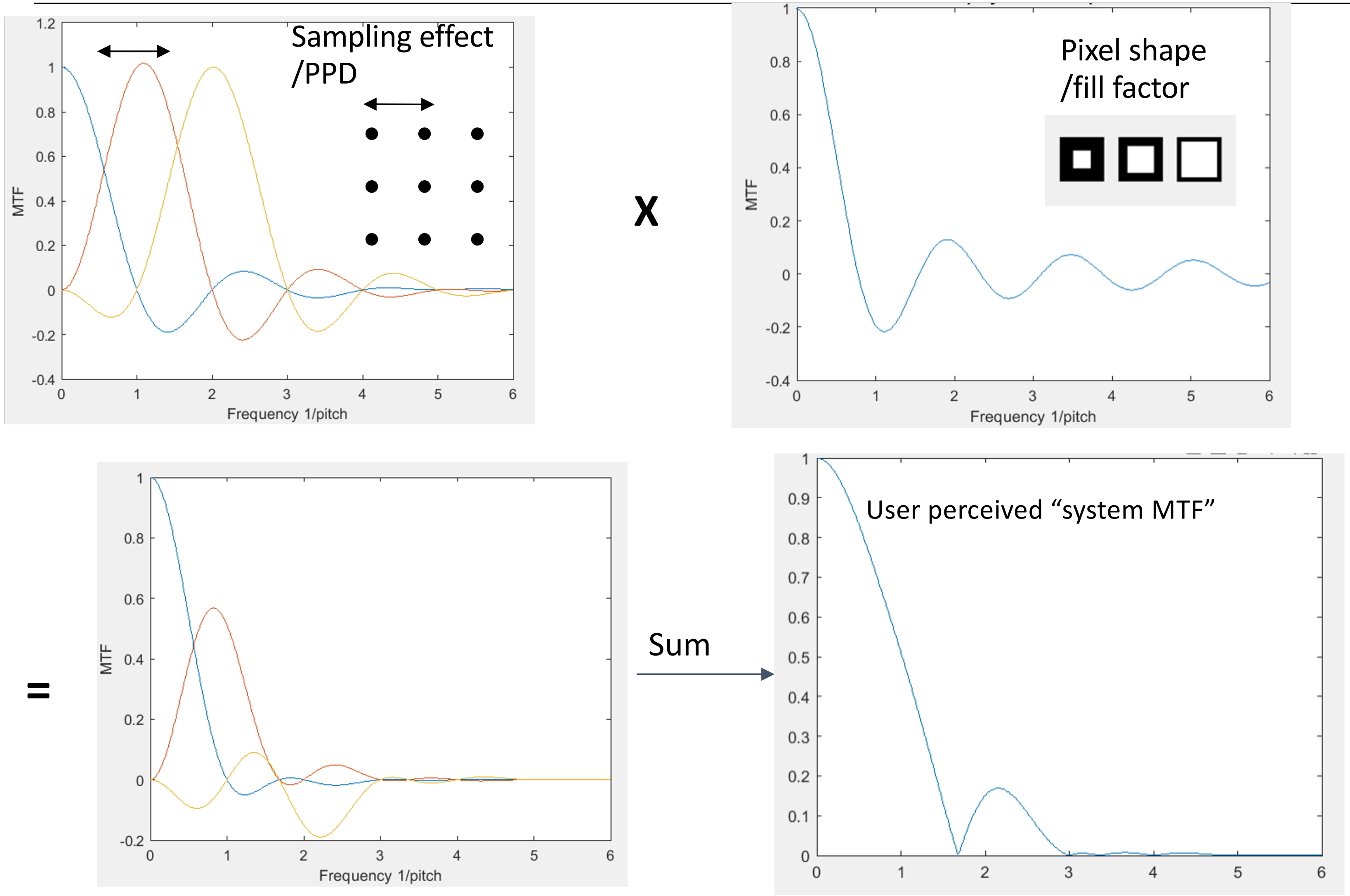
Fill Factor is not monotonic @low PPD





- **Key takeaway**: when display resolution is low (around 20-30 PPD), there is an optimal FF (not the higher the better). “System MTF” method can find it.

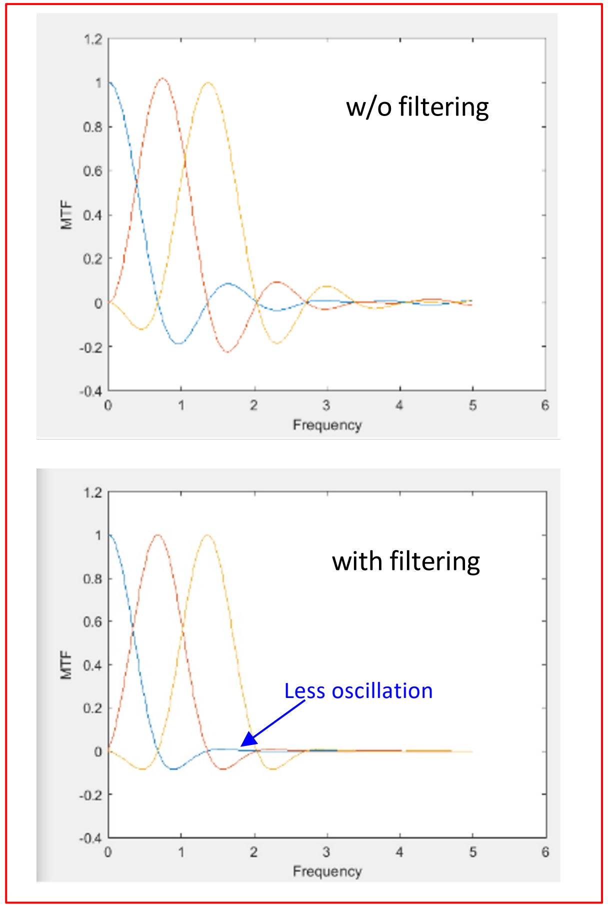
Unify PPD and fill factor in *frequency domain*

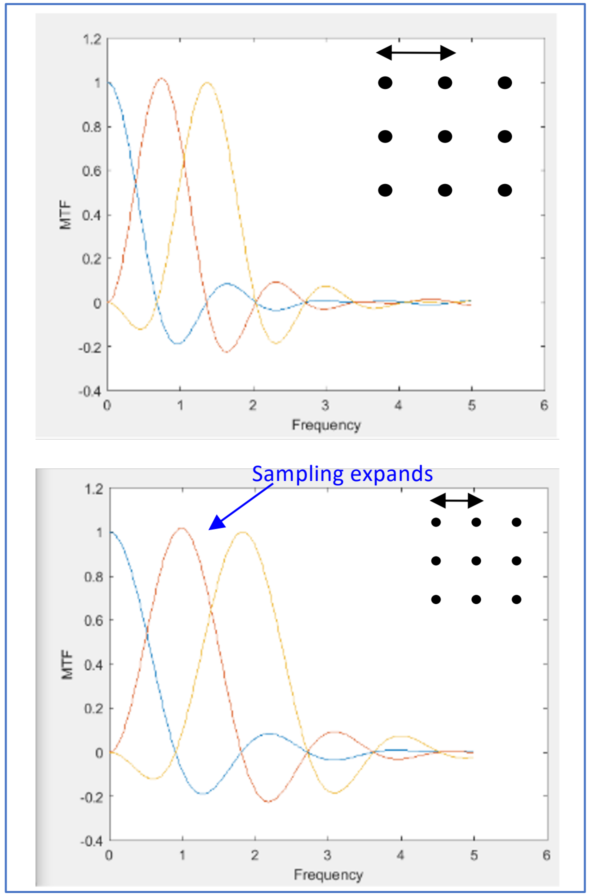


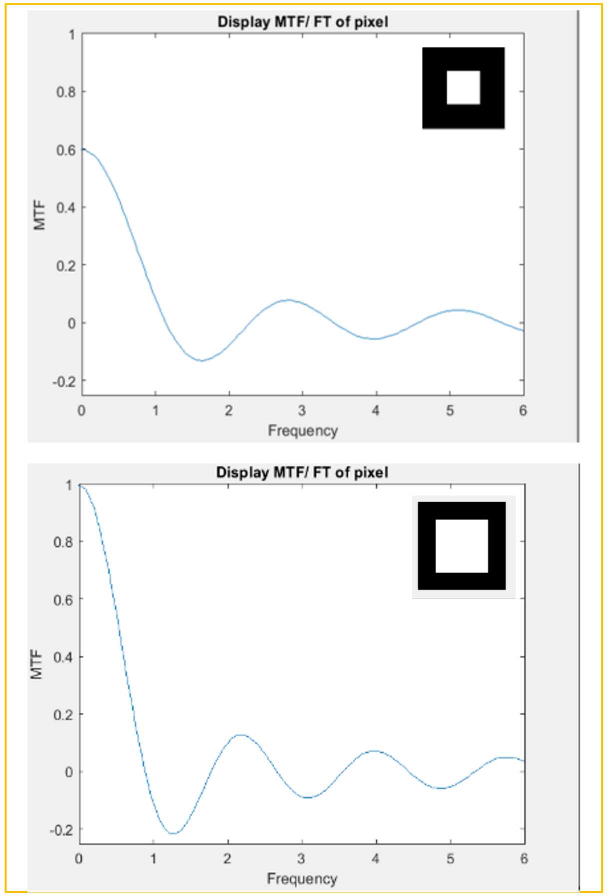
Key message:

Fill factor and pixel shape both have a strong impact on retina display experience. We cannot just increase PPD without optimizing fill factor.

Quantify the effect of (PPD, Fill factor) in frequency domain:





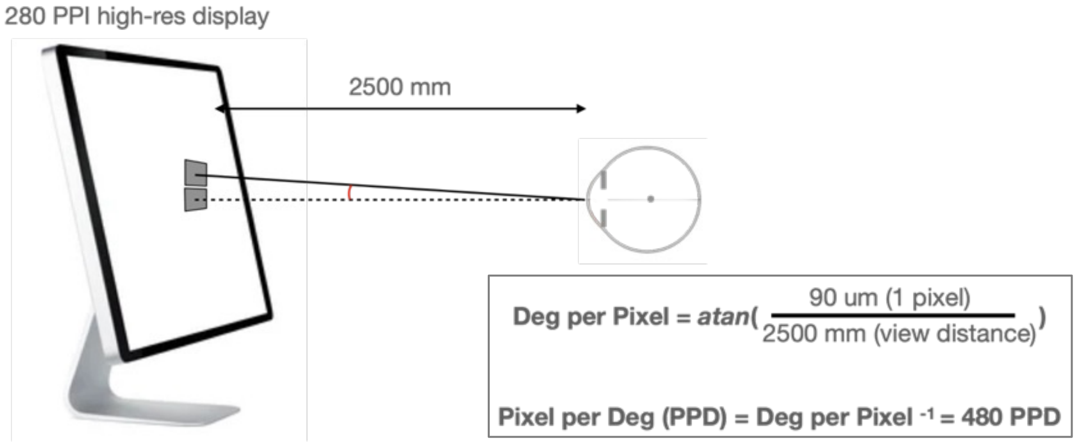


$$SystemMTF = \int_0^{40}\int_0^{\pi}\frac{1}{p}\int_{-p/2}^{p/2}[(H_{gh}(\xi,\theta)e^{-j2\pi\xi t}) * \sum_n \delta(\xi - n\frac{1}{p(\theta)}) \times PIX(\xi,\theta)]H_{opt}(\xi)w(\theta)w(\xi)dtd\theta d\xi$$

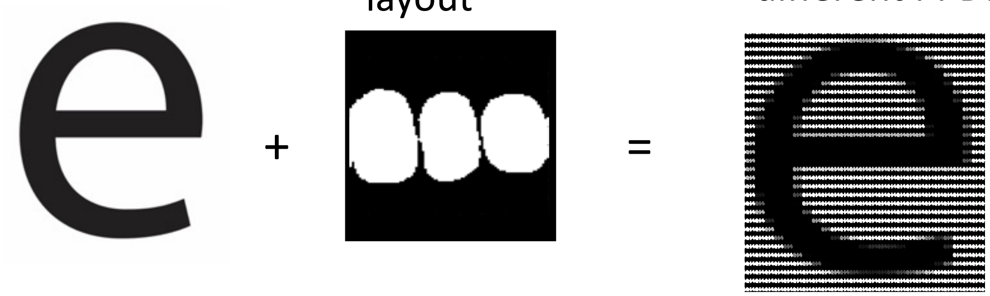
Linking perception and math model

User study gives “system MTF” perceptual meaning

Setup



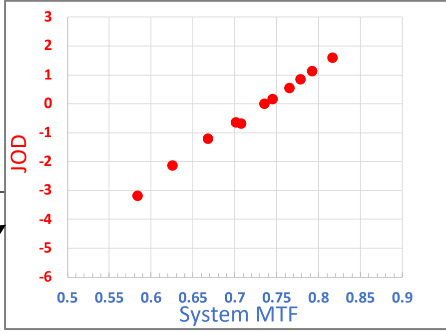
Content + Display pixel layout = Emulated different PPDs

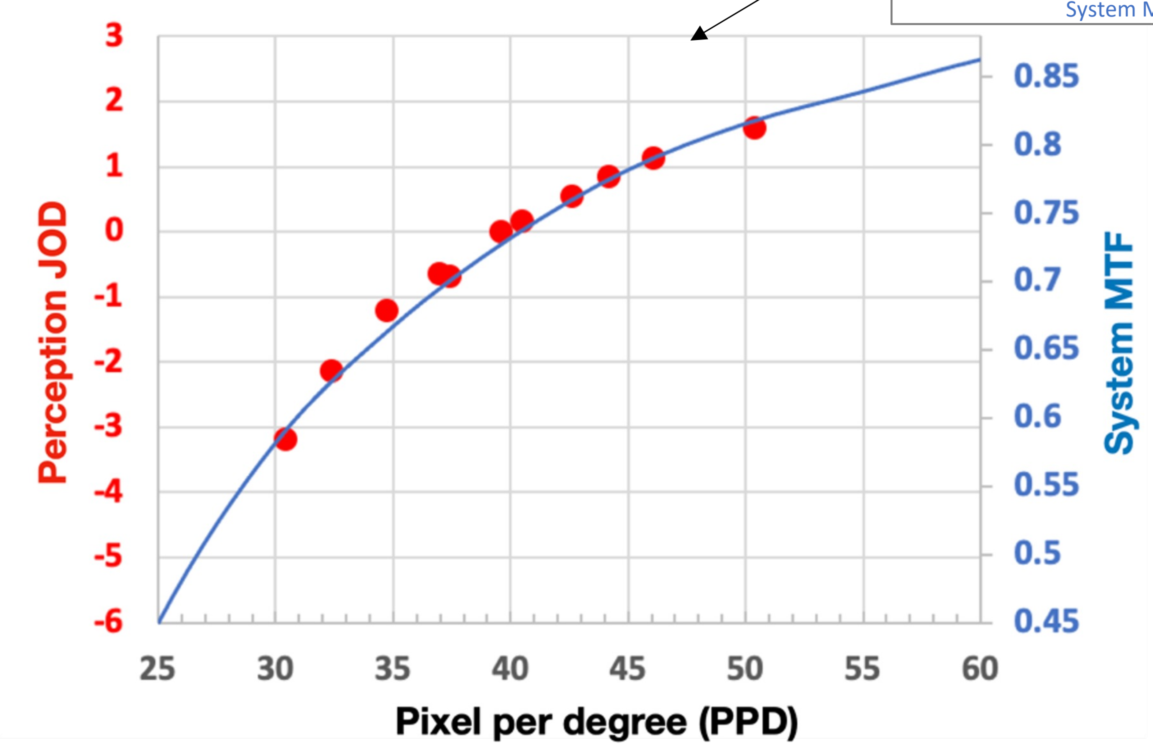


- User study is done by emulating low-PPD displays by a very high-PPD monitor (e.g. each virtual pixel is 10x10 to 20x20 physical pixels)

Result

Same data





- System MTF matches user data nicely, validating the metric
- Human perception is very sensitive to see 0.05 MTF change (~1 JOD)!
- 25 to 30 PPD is a big jump (3 JOD); 50 to 60 PPD is still a meaningful improvement (1 JOD)

Applications of the model

Quantify effect of display rotation by 21 deg (demo)

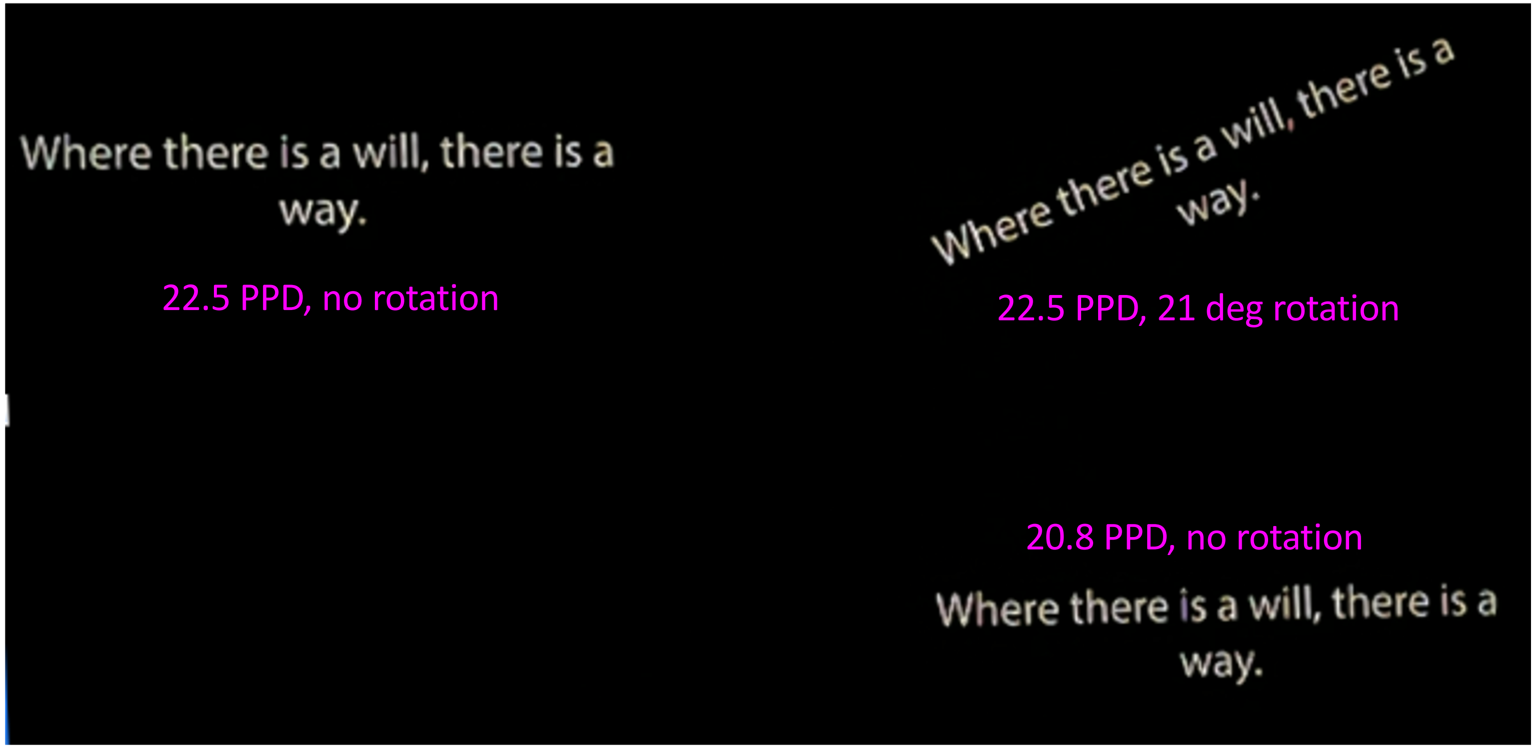
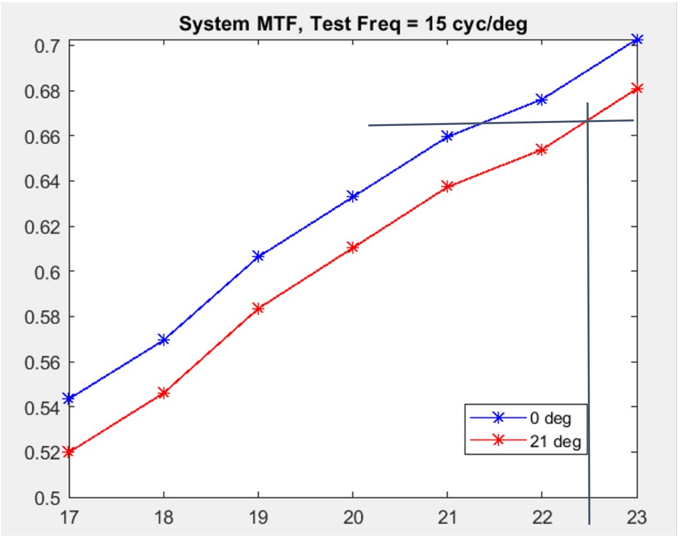


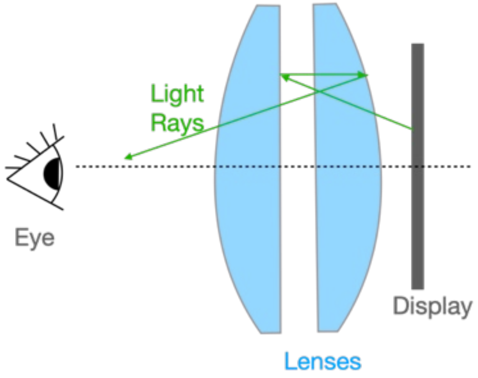
Photo capture of emulation on a high-res display



- **Key takeaway**: Display rotation reduces the effective PPD and the weight function over angle is critical. For Arcata, 22.5 PPD becomes 20.8 PPD for vertical-dominant content (English text)

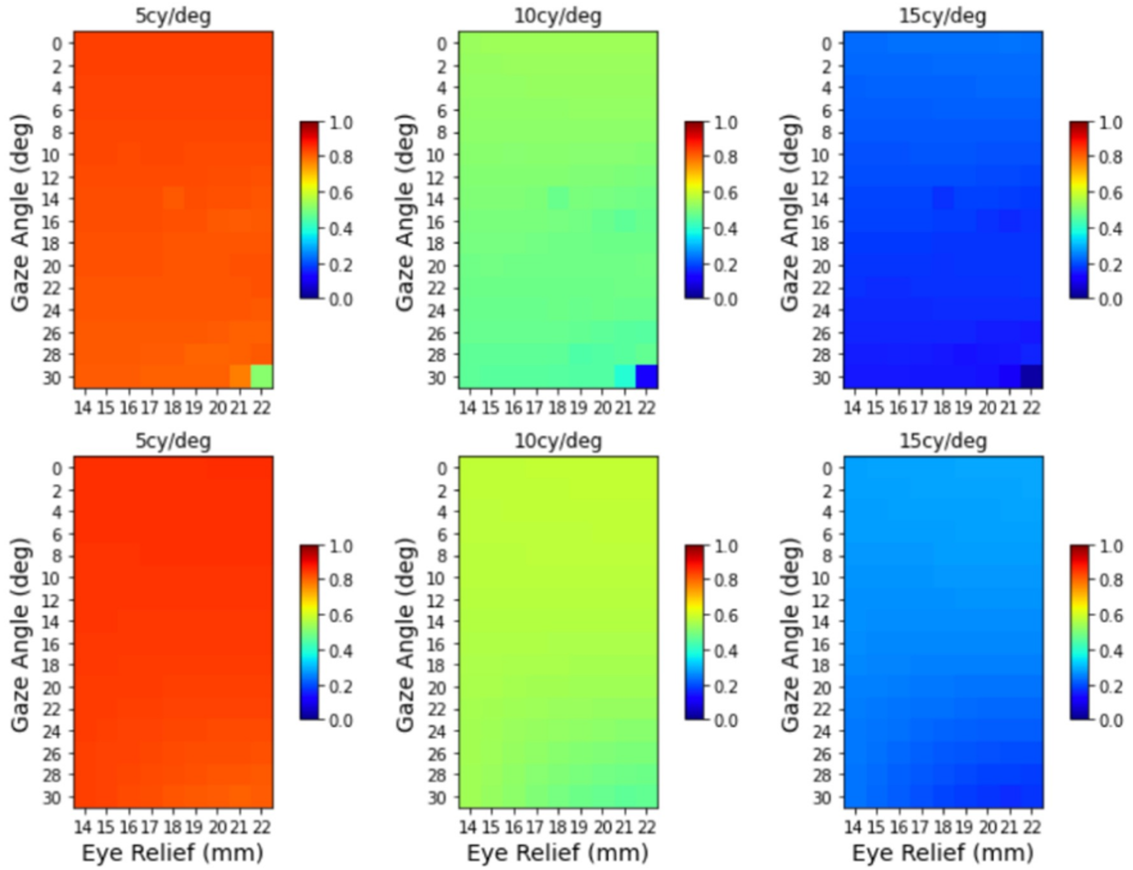
Evaluate optical design and architecture: it’s not just PPD

- Correctly predicted similar performance in Design1 (54PPD, a lot more challenging) and Design2 (36PPD, more practical)
- Must evaluate Lens + Display combined MTF (lens-only MTF is a weak predictor)



VR System Design 1: 36PPD
with high fill factor

VR System Design 2: 54PPD
with low fill factor



System MTF: higher is better

Optimal graphical filter given a display hardware

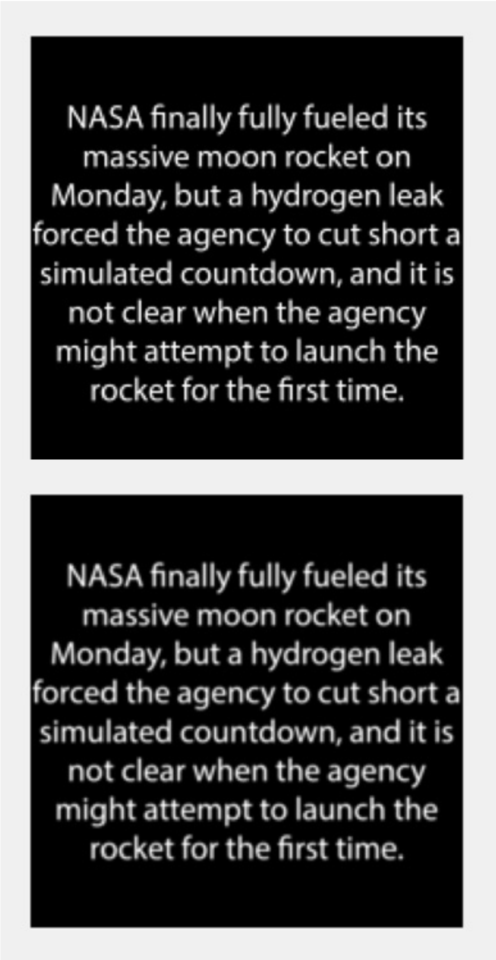
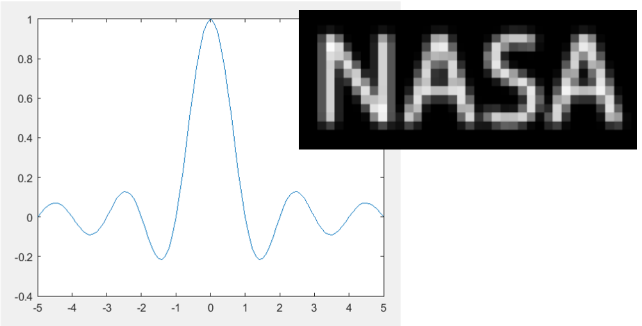
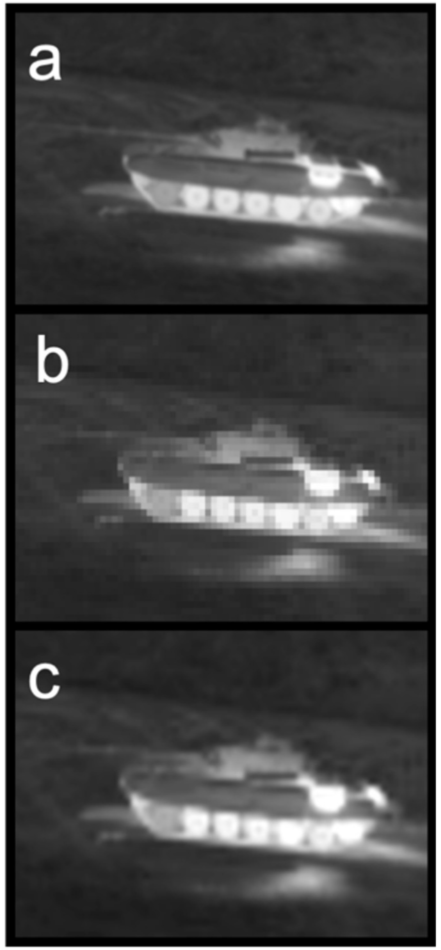


Photo capture of emulation on a high-res display

Simple down sampling

With special graphical "filtering" (same HW)





Source

Simple down sampling

With special "filtering"

- **Key takeaway**: For a given display HW, proper image processing on source graphics can maximize the user experience. "System MTF" provides the quantified method for this type of optimization.

Optimization on system level

Source image/ content (e.g. 2000x2000)

↓ ← Pre-blur /graphics pipeline

Down sampled image before display (500x500)

↓ ← Post-blur /filtering

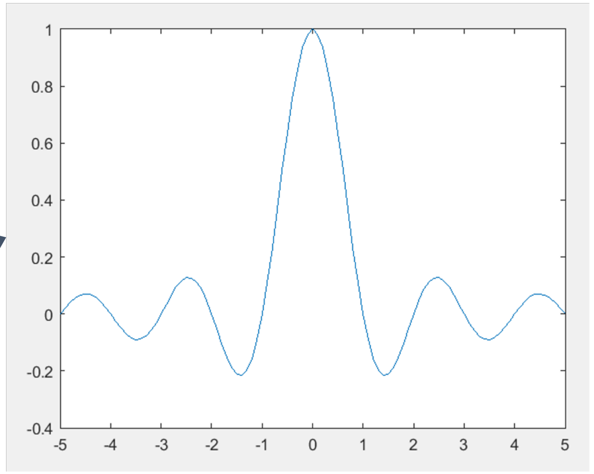
Image shown on display (to the user) (500x500) ← Display impact: PPD/ Fill factor

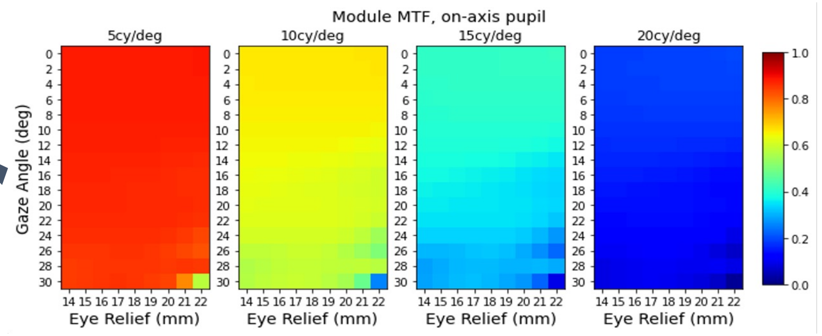
↓

Optics + Eye blur ← Optical design

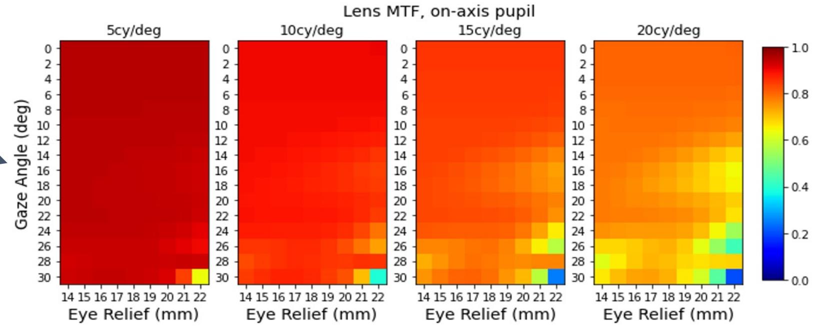
↓

Retina + brain





Module MTF



Lens MTF

Conclusion

A complete process to define and deliver user experience

1 Experience metric	2 Perception/ vision metric & spec	3 HW design & tolerance against engineering metric	4 HW engineering: process, integration, calibration/test
What matters to typical users	*Need to be translatable to executable engineering spec*	*Design vs. Spec: can this be manufactured with tolerance?*	*New processes/materials; calibration; factory*
"Good sharpness targeting for photographic/ text quality"	• Developed "system MTF" model to capture many contributors, match user experience, and guide practical product design • Targeted content determines frequency	• Display pixel density/ PPD • Pixel Layout, fill factor • Lens /Optics MTF • Graphics • Combined camera MTF (for MR)	• New displays to maximize system MTF? • Test methodology? • Don't over-engineer/-spec optics