SPECIAL SECTION PAPER

WILEY

Local pupil swim in Virtual- and Augmented-Reality: Root cause and perception model

Jerry Jia, SID Member ⬤ | **Tsz Tai Chan, SID Member** |
Trisha Lian, SID Member | **Kevin W. Rio, SID Member**

Meta Reality Labs, 1 Hacker Way, Menlo Park, California 94025, USA

Correspondence

Jerry Jia, Meta Reality Labs, 1 Hacker Way, Menlo Park, Burlingame, CA, 94025, USA.
Email: jerry.jia@meta.com

Abstract

The optics between display and human eye in a typical VR/AR headmounted display (HMD) can introduce a common visual defect - local pupil swim (also called local ripples or "orange peel" effect), where virtual content distorts locally with head movement. Compact optical design (such as pancake optics) is increasingly sensitive in design and manufacturing tolerance to this perceptual effect. This work provides a method to root cause and quantify the impact based on perceptual modeling, optics simulation, and measurement.

KEYWORDS

distortion, perception, pupil swim, visual ripples, VR

1 | INTRODUCTION

Typical virtual reality head-mounted display system (VR HMD) has optical elements between display and human eye. The design sensitivity and manufacturing tolerance of these optical elements and surfaces can generate a visual defect of local ripples. An illustration of this type of defect is shown in Figure 1. The two bands' regions (between the dotted red lines and between the dotted blue lines) have unnatural distortions. This defect is particularly obvious with these observation conditions: (1) when the virtual content is world-locked, meaning the display renders content dynamically with head motion to maintain a stable virtual world; (2) when people are conducting an eye motion called vestibulo-ocular reflex (VOR), where gaze is fixated on a virtual object (e.g., a letter on a wall of text) while head is moving. This is a frequent use mode in VR, which maintains users' visual continuity despite head movement. This is in contrast with eye's saccade where users lose visual continuity due to temporary blindness during saccade; (3) when the virtual content is highly textured or of high-spatial frequency (e.g., text and find grid); (4) when the visual defect is projected to the fovea region.[1,2] This visual defect is often referred to as "orange

peel" effect due to its resemblance to the skin of an orange. In our work, we name the perceptual effect of the dynamic distortion as local pupil swim (in contrast with global pupil swim,[3] more in Section 4.1).

A recent popular VR architecture is the pancake optics (Figure 2A).[4-6] Compared to the older singlet optics design (only refractive lens, similar to a magnifier lens placed in front of a display), pancake optics has a folded and reflective optical path which can potentially make the local pupil swim more severe and sensitive to manufacturing tolerance. This phenomenon also exists in AR system (Figure 2B) where multiple layers of optical stacks (e.g., lamination layers, coatings, lenses, prescription correction, and eye tracker) can introduce a similar visual defect.

Interestingly, the human perception system is not as sensitive when the distortion error is presented statically (e.g., no head or eye motion) as when the defect is observed dynamically with VOR. Understanding the root cause and quantifying the perceptual effect of local pupil swim in different observation modes are essential for its mitigation in design, product architecture, and manufacturing. Correlation of the human perception to engineering metrics (such as optical surface quality) provides the foundation for making quality products.

FIGURE 1 Illustration of local pupil swim experience (not data from a physical sample). The experience manifests as local ripples when people conduct VOR in VR/AR headsets with world-locked content. The regions between the dotted lines have unnatural distortions.
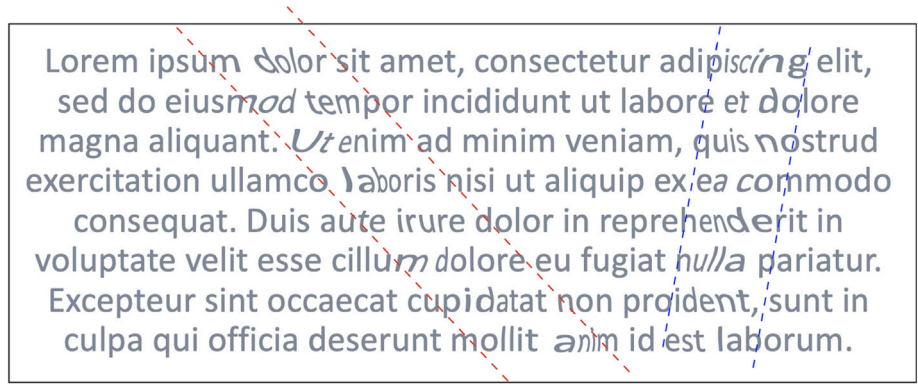

FIGURE 2 (A) Typical schematics (not ray tracing) of pancake optics used in VR HMD. Compared to singlet optics, the folded reflective path in pancake architecture can potentially be more sensitive for having local pupil swim visual defect. (B) A typical structure for an AR system. The multilayer optical stack in front of the eye can introduce similar local pupil swim as in VR.
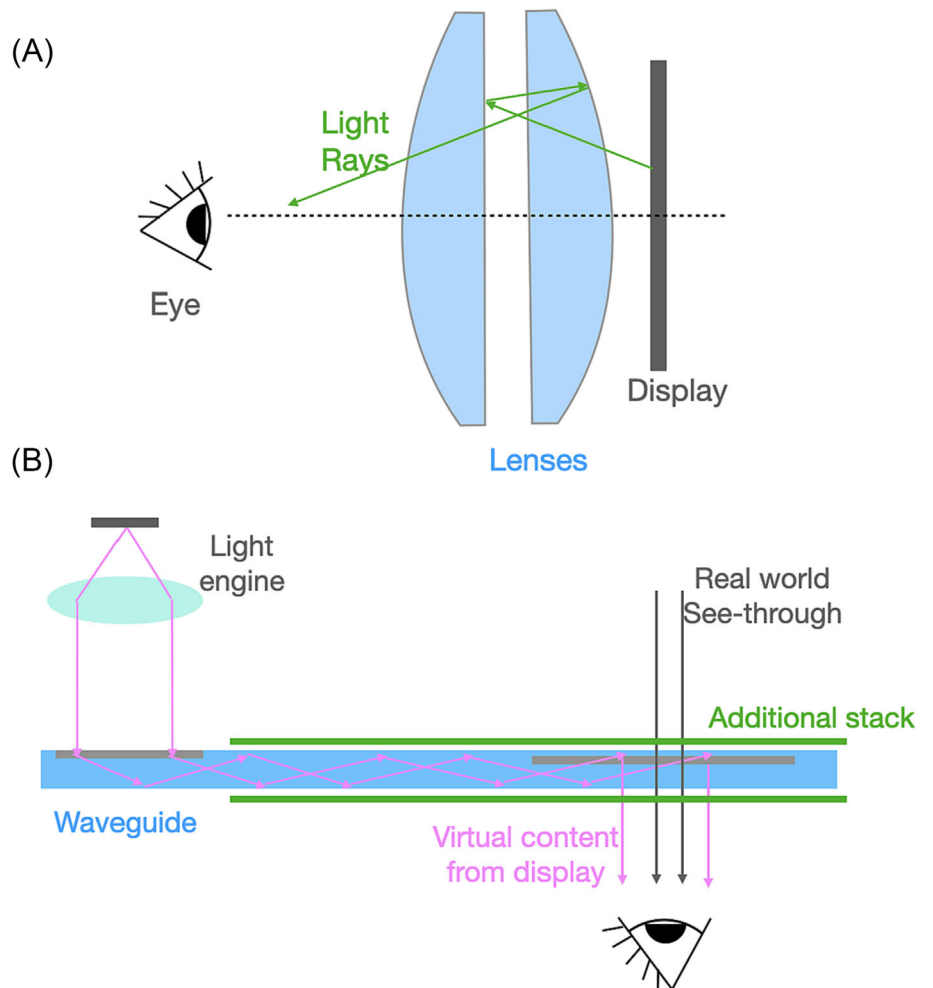
2 | ROOT CAUSE OF LOCAL PUPIL SWIM

By design, a perfectly made optical system does not have local pupil swim (mid-spatial-frequency waviness). When there is a small defect on the lens surface, such as a crack or bump, it introduces local distortion at the corresponding field angle location. When a user fixates on a virtual object and moves their head, that is, performing head motion and rotating the eyes in the opposite direction during VOR, the distortion changes locally since the defect is moving to another location in their visual field of view (FOV), as illustrated in Figure 3. This can be very noticeable and annoying since the virtual
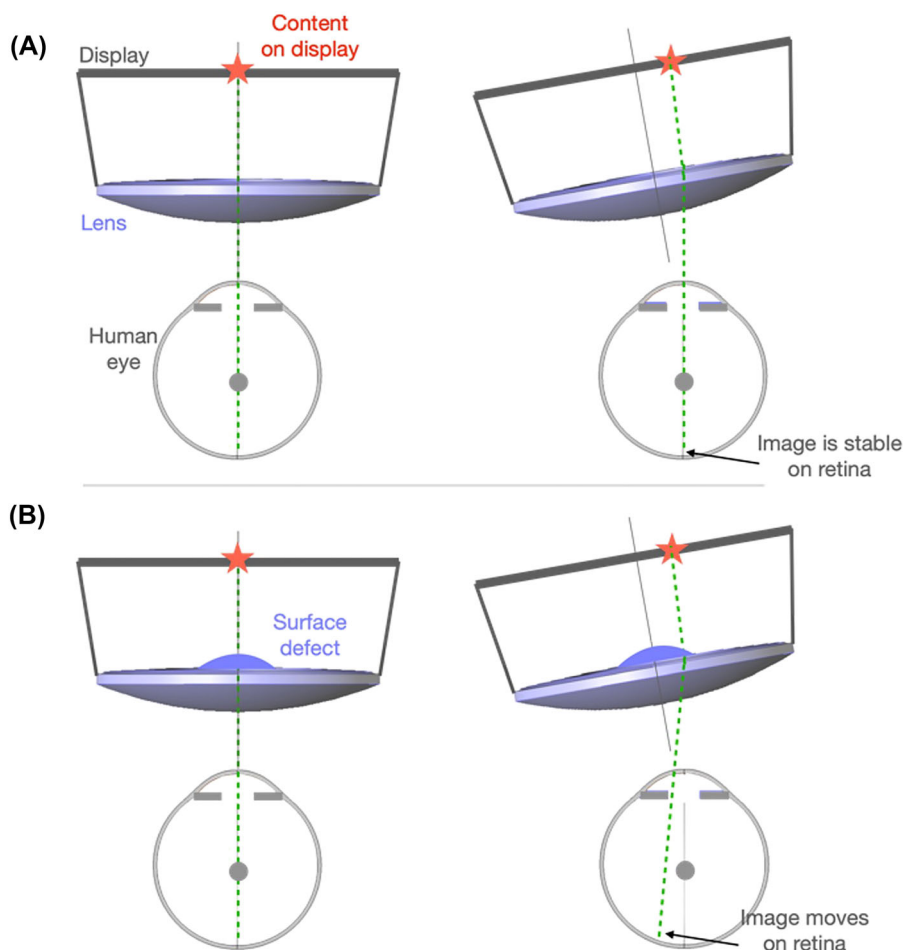
world is supposed to be static. In the VR headset, VOR with a slow head rotation speed (e.g., 5–10°/s) and a magnitude of +/− about 10° is best to observe the local pupil swim. Typically, a complete local pupil swim defect is >5° wide (e.g., the bands in Figure 1). So the head needs to move wider than defects' angular width to confirm the perceptual defect with a full cycle.

We dive deeper in the following sections but here are some high-level insights:

1. The perceptual outcome depends on the ripples' magnitude and size of surface defects
2. Human perception system is very sensitive to local pupil swim during VOR (even magnitude of 0.01° is easily noticeable)
3. Optical system design, location of the surface in the system, its function (e.g., reflection and refraction), and defects' location on the surface all impact the perception of local pupil swim.

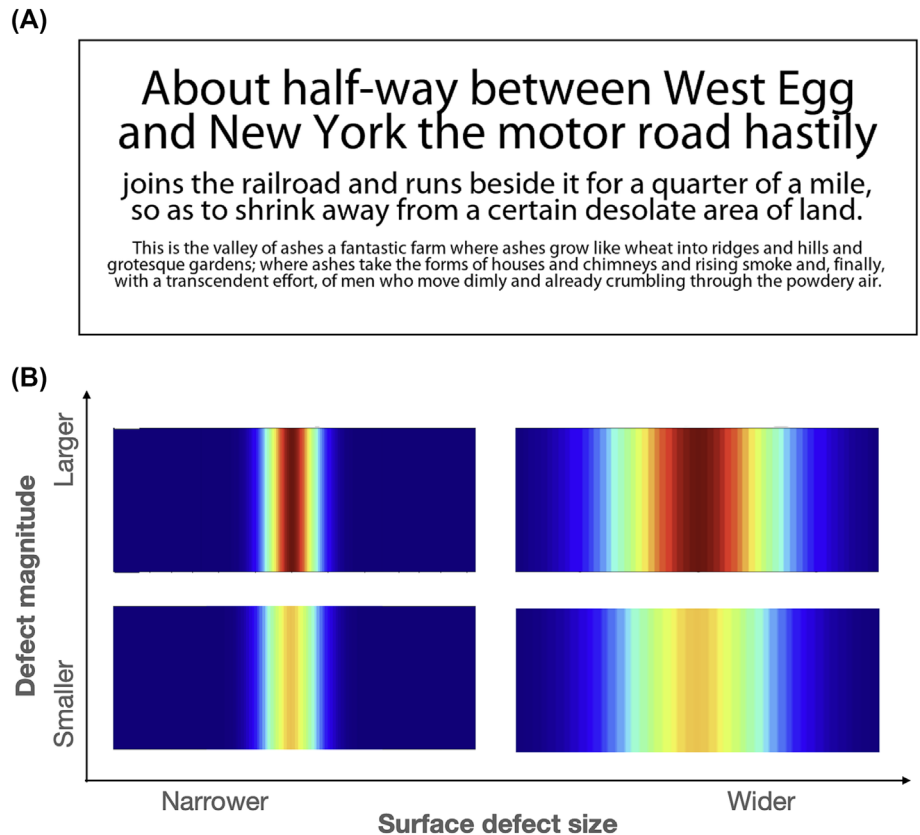This work aims to quantify the impact of local ripples and provide a method to mathematically link the characteristics of defects on both component level and system level to a perceptual metric.

3 | PERCEPTION MODEL AND USER STUDY

3.1 | Method for the user study

We used text of different font sizes as the content for the user study (Figure 4A) because we found that regularly textured content (with mid- and high-spatial frequencies) was most revealing for local pupil swim. We then overlaid local distortion defect (in angular space) on top of this content and generated a dynamic video to show on a retina resolution display (MacBookPro 16″). The content stayed static except for the distortion effect (similar to Figure 1 but the distortion defect bands move horizontally across the content). Users kept their head and eyes steady during the study. We confirmed this experience was very similar to that of VOR in a VR/AR headset with world-locked content.

FIGURE 4 (A) Text content used in the user study. (B) Illustrative examples of surface defects with varied sizes and amplitudes. Optical simulation converts the surface defect into distortion in angular space; then, it is overlaid on top of the text content to generate a dynamic video.

(A)

About half-way between West Egg and New York the motor road hastily joins the railroad and runs beside it for a quarter of a mile, so as to shrink away from a certain desolate area of land. This is the valley of ashes a fantastic farm where ashes grow like wheat into ridges and hills and grotesque gardens; where ashes take the forms of houses and chimneys and rising smoke and, finally, with a transcendent effort, of men who move dimly and already crumbling through the powdery air.

(B)

We generated artificial surface defects with Gaussian shape to model the defect magnitude and size along the horizontal axis. Figure 4B showed a few examples of these surface defects. They could be seen as a vertical bump on the lens surface. In optics, surface errors are usually represented by slope error (first derivative of surface profile), so the defect magnitude is defined as the peak slope error (not the amplitude of the Gaussian) and the size is defined as 4 sigma of the Gaussian function. This is just a convention of definition: integrating slope and size is equal to the amplitude of Gaussian.

To convert the surface defects (on lens surface with the unit of µm/mm) into what the eye sees as local pupil swim (in angular space with the unit of degree), we performed optical ray tracing for each condition in Zemax OpticStudio. For the user study, we tested defects with magnitude ranging from 0.1 to 3 µm/mm and sizes from 0.3 to 6 mm. The effect of local pupil swim from distortion can be emulated by applying the distortion with varying spatial correspondence in the FOV to an image. We modulated the distortion based on a simulated head motion profile to create videos. After generating the videos, we presented them in random order to subjects and asked for their perceptual rating in a 5-point ITU-R quality scale[7] in Figure 5.

Scale	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

FIGURE 5 A 5-point scale (ITU-R quality scale[7]) for subjects to rate the local pupil swim.

3.2 | Result of the user study

User study result is shown in Figure 6. The x axis is the surface defect size, and the y axis is the surface slope error. The colored contour lines (green to blue gradient) are the average scores of users reported scales according to Figure 5. The average standard deviation among 12× subjects for all test conditions is 38%. The two black lines labeled with 0.01° and 0.02° are the peak magnitude of local pupil swim in degrees corresponding to the conditions defined by surface error slope and surface defect size. The local pupil swim characteristics are obtained from ray tracing simulation using the same surface defect conditions.
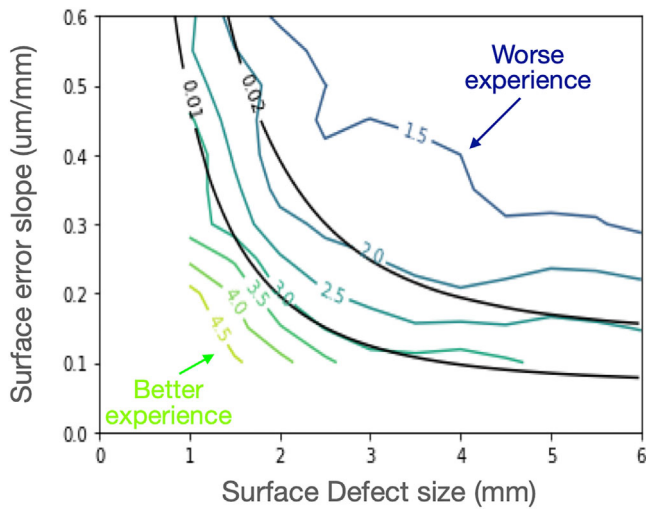
FIGURE 6 User perceptual ratings (the colored contour lines) at different defect magnitude and size. Smaller value represents worse experience (as defined in Figure 5). "0.01°" and "0.02°" black lines are the contour lines of local pupil swim magnitude of 0.01° and 0.02°, respectively, with the corresponding surface defect size (x axis) and slope (y axis).

We found that users are generally more tolerable to very small-sized defects (<1 mm, as shown on the left side of the x axis). The less noticeable small-sized defects can be partially explained by the optical smoothing effect: The light beam going through the lenses' surfaces typically has the size of 1–4 mm in diameter as defined by both the eye's pupil size and optics design. This beam interacts and convolutes with the small defects, and the distortion is averaged out to be small.

Another important finding from the perceptual score contours in Figure 6 is that there is a predictable relationship between surface defect size and surface error slope on the same contour line of user score. Further this relationship strongly aligns with local pupil swim magnitude (e.g., the 0.01° and 0.02° curves). In other words, it is likely that surface error slope and size can be combined into one single metric which is proportional to the peak local pupil swim (see the next section).

3.3 | Perception model linking user experience and engineering metrics

We have explored the following three things: the component surface error (magnitude in μm/mm and size in mm), the system-level metric local pupil swim (with the unit of degree), and the perceptual score (defined in Figure 5). Our goal is to correlate all three using simple formulas. The basic hypotheses/desirable laws are as follows:
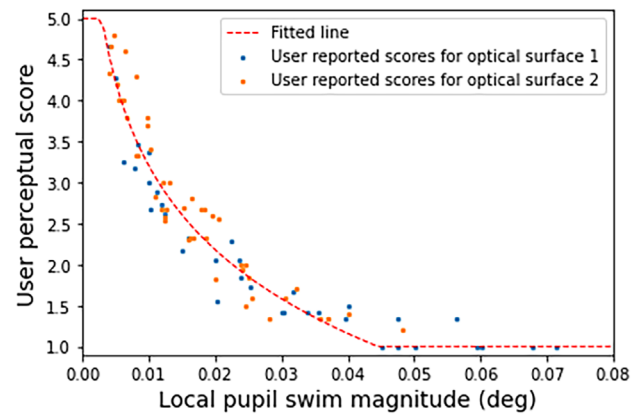

FIGURE 7 The correlation between peak local pupil swim (x axis) and user perceptual score (y axis). The dots are with different conditions (surface defect size, magnitude, and optical surface in the system) containing the data in Figure 6. The two colors of the dots are from surface defects on different surfaces in the system. The fitted line is based on Equation (1).

1. Local pupil swim is the direct root cause of the perceptual complaint score, and they are proportional to each other with simple relationship (linearly or logarithmically).
2. Local pupil swim is also physically linked to surface defect, and the defect size and magnitude are ideally combined into a single metric using their relationship revealed in Figure 6.

To achieve (1), we plotted the same data of Figure 6 in Figure 7. The local pupil swim peak magnitude (x axis in Figure 7) is correlated with user perceptual score (y axis). The correlation can be captured by Equation (1).

Combining the optical simulation and the user study results, we have the local pupil swim P_{local} from component surface error as system metric (with the unit of degree) and the user perceptual quality. We linked two metrics together by the Weber–Fechner's law[8,9]:

$$R = k\,ln(P_{local}/P_0). \qquad (1)$$

R is the user perceptual score, k is a constant, P_{local} is the peak magnitude of in the local pupil swim map, and P_0 is the threshold of perceiving the local pupil swim. We optimized the parameters in Equation (1) and obtained the fitted function (red dash line in Figure 7) with mean-absolute error of 0.211 and R^2 value of 0.932. Both ends for user scores (1 and 5) in the function were clipped since the users were asked to rate from 1 to 5 according to the 5-point scale in Figure 5. From the user study, we tested defects with a range of size and magnitude on two different optical surfaces. This correlation holds true

independent of surface defect size, magnitude, and location of the optical surface in the system.

To achieve (2)—linking local pupil swim to surface defects and combining defects' characteristics into a single metric—it takes an iterative simulation and experiment approach. In this section, we will give the high-level concept and result. In the next section, we provide additional details.

Given the different magnitudes and sizes of defect, we wanted to find a single metric that can correlate directly with the local pupil swim magnitude. Based on the relationship revealed in Figure 6 and optical simulation, we found that this correlation can be achieved by applying a 2D spatial convolution filter to the surface slope error:

$$S_{conv} = S * f_l, \tag{2}$$

where S is the 2D map of surface error slope; f_l is a 2D spatial convolution filter, whose setting is dependent on the optical design; and S_{conv} is the convoluted surface slope error map.

We simulated defects on two optical systems with different optical surfaces and performed ray tracing to get the local pupil swim magnitude. After optimizing the parameters in Equation (2) for each optical surface, the convoluted surface slope error is approximately proportional to the local pupil swim magnitude as shown in Figure 8.

Therefore, the local pupil swim can be represented by

$$P_{local} = w_l S_{conv} = w_l (S * f_l), \tag{3}$$

where P_{local} is the local pupil swim map (in degrees) and w_l is a weighting coefficient depending on optical design. The weighting coefficient can be interpreted as the surface sensitivity to defect, where a higher number means that surface is more sensitive to defects and thus leading to tighter tolerance needed and increased difficulty of manufacturing. An example is shown in Figure 8. Optical system A's green curve has a much higher sensitivity to local pupil swim than optical system B's green curve, that is, with the same convoluted surface error (e.g., 0.1 μm/mm), system A has higher local pupil swim/ripples than B.

4 | VALIDATION BY SIMULATION AND METROLOGY

4.1 | Measurement of local pupil swim and surface error

To calibrate and validate the model and metrics above, real samples' pupil swim data and component surface error data were collected. Lens surface error is typically measured by commercially available interferometer (more often used on flat surfaces, but also on curved surfaces) and profilometer (for both flat and curve surfaces, e.g., Panasonic UA3P). A dense 3D point cloud can be obtained from these methods and can be input into Zemax optical simulation[10] or other analysis programs. Section 4.2 details how we process these raw data with our special smoothing algorithm to gauge against the component specifications.

To measure pupil swim, we designed a wide-angle fish-eye type camera (a conoscope with FOV of $+/-75°$ and resolution of 66 pixel per degree) with similar aperture sizes matching human eye (Figure 9A). This camera system is integrated into an optical alignment system with additional functions to form a manufacturing-friendly calibration/test station. Other options exist but we used a pre-distorted dot pattern (single pixel, 0.5° or
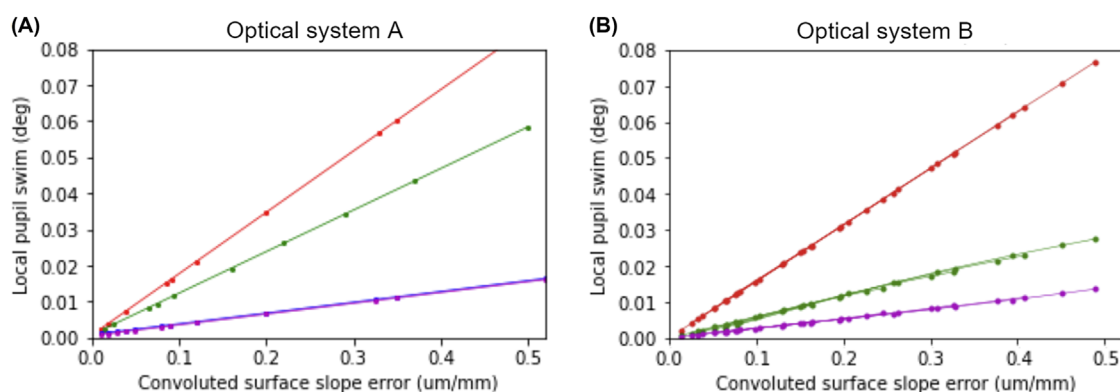

FIGURE 8 Local pupil swim magnitude against convoluted surface slope error from two different optical systems (simulation; see Section 4 for experiments). Each color represents a different optical surface in the optical systems. A key difference between the two systems is the sensitivity of the green curve.

1° distance apart) to form a supposedly regular angular space (e.g., 1° step within +/−50° FOV). Due to the surface error within the optical system, the exact angular location of each dot is not on the regular grid and its difference can be plotted as a vector field in 2D angular space (Figure 9B). In the manufacturing implementation, we have practically consider other factors such as the fact that display pixels are not exactly aligned with angular steps (e.g., cannot display a signal at a non-integer pixel location) and the extreme accuracy of camera calibration to map pixels to angles. We do not get into those details in this report.

Pupil swim is essentially the difference between two angular distortion maps at two pupil locations.[3] For example, we initially foveate at (0,0)° in the FOV, then we perform VOR and rotate the head by 1°. During this process, the system pupil (also eye pupil) shifts from on the optical axis to off the optical axis. We can take the
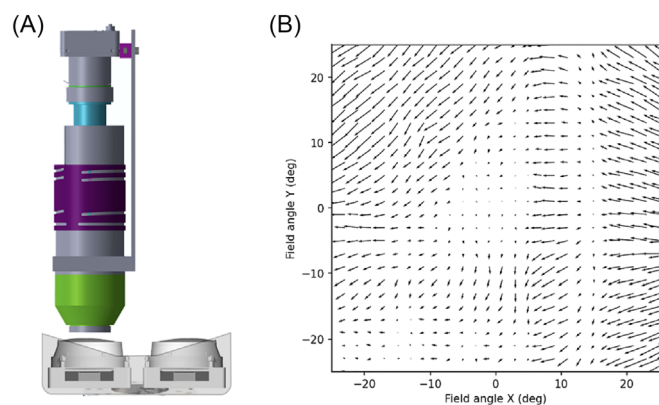
difference of the two angular spaces and plot this difference as a heat map shown in Figure 10A. These data contain two sources of pupil swim: (1) global pupil swim that is caused by optical design itself. When the system pupil shifts around the optical axis, the distortion behavior of the system changes causing global pupil swim. This portion of pupil swim is captured in Figure 10B. (2) Local pupil swim that is caused by lens surface errors. While the magnitude of (1) is typically much larger than (2), human perception is a lot higher for local pupil swim. Global pupil swim tends to have a slower spatial frequency than local pupil swim. When we observe the visual content through the lenses, particularly with small head and eye rotation (e.g., <10°), we don't always sense the global pupil swims, this is because the sensitivity threshold of detecting the pupil swim depends on their spatial frequency.[11] The user study conducted in this work and simulations have been solely focused on local pupil swim.

To evaluate local pupil swim, we subtract the global pupil swim (Figure 10B) from the measured pupil swim (Figure 10A) to obtain the local pupil swim as shown in Figure 10C. This local pupil swim is much better correlated with perceptual result than as-measured pupil swim.

One practical technique we applied in manufacturing tests is that we did not move the camera or the sample during testing. Since the global pupil swim is due to location change of the system pupil/stop, we can approximate the local pupil swim result by subtracting an angular map by its virtually rotated copy. Considering the VOR direction and magnitude can be arbitrary in real usage, we determine the local pupil swim by a weighted combination of numerous VOR cases (e.g., 1°, 2°, and 5° VOR angles).
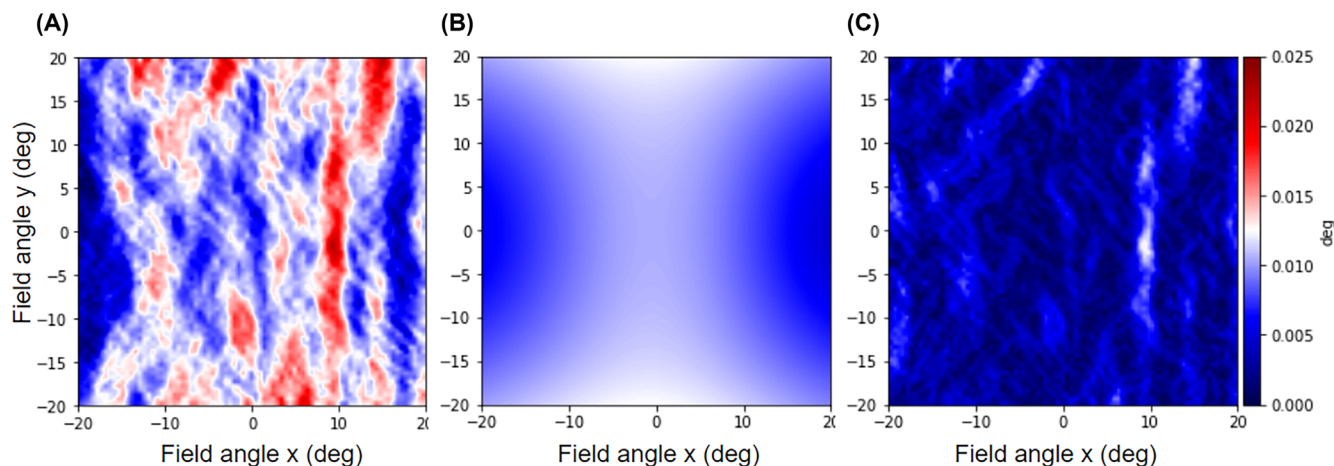

FIGURE 9 (A) Conoscope camera as part of the test system to measure pupil swim. (B) An example of distortion error map captured by the camera.

FIGURE 10 (A) Pupil swim measured with VOR rotation of 1°. (B) Global pupil swim originated from optical design (from simulation). (C) Measured pupil swim resulted from (A) subtracting (B).

4.2 | Kernel smoothing and optical simulation approach

In a VR or AR system, light beam originated from the display passes through or reflects off different optical surfaces. An example of a VR optics design is shown in Figure 11. The size of the beam grows as it propagates through the surfaces and ultimately hits the eye (top schematics in Figure 11). The beam size is determined by the eye pupil size, optical design, and the propagation location. Pupil size typically ranges from 2 to 7 mm in diameter depending on the illumination of the visual content.[12,13] In the VR and AR system, the eye pupil is also the system optical stop.[14] This means when the pupil size is large, the light from the same spot on the display interacts with a larger area of the lenses' surface. When the defect is small, less portion of the light beam is affected by the defect, so the resulting distortion is smoothed out and has a small magnitude. We select pupil sizes of 3 or 4 mm. We also explored the effects of pupil size on the relationship between local pupil swim size in degree and defect size in mm as shown in the bottom plot of Figure 11.

For component surface error, we first measured the surface error map S by a profilometer or interferometer. The process is relatively standard. Then, S_{conv} is obtained based on Equation (2). The choice of the convolution filter f_l is based on optical design, pupil size setting (3 or 4 mm), and location of the optical surface as stated previously. Figure 12 showed an example of raw measured slope error map and the corresponding convoluted results. By the same principle applied to the measured pupil swim, we processed the slope error map to get the local pupil swim features from the surface defect. Figure 13A showed the resulting surface slope error, and analysis will be done in the next section.
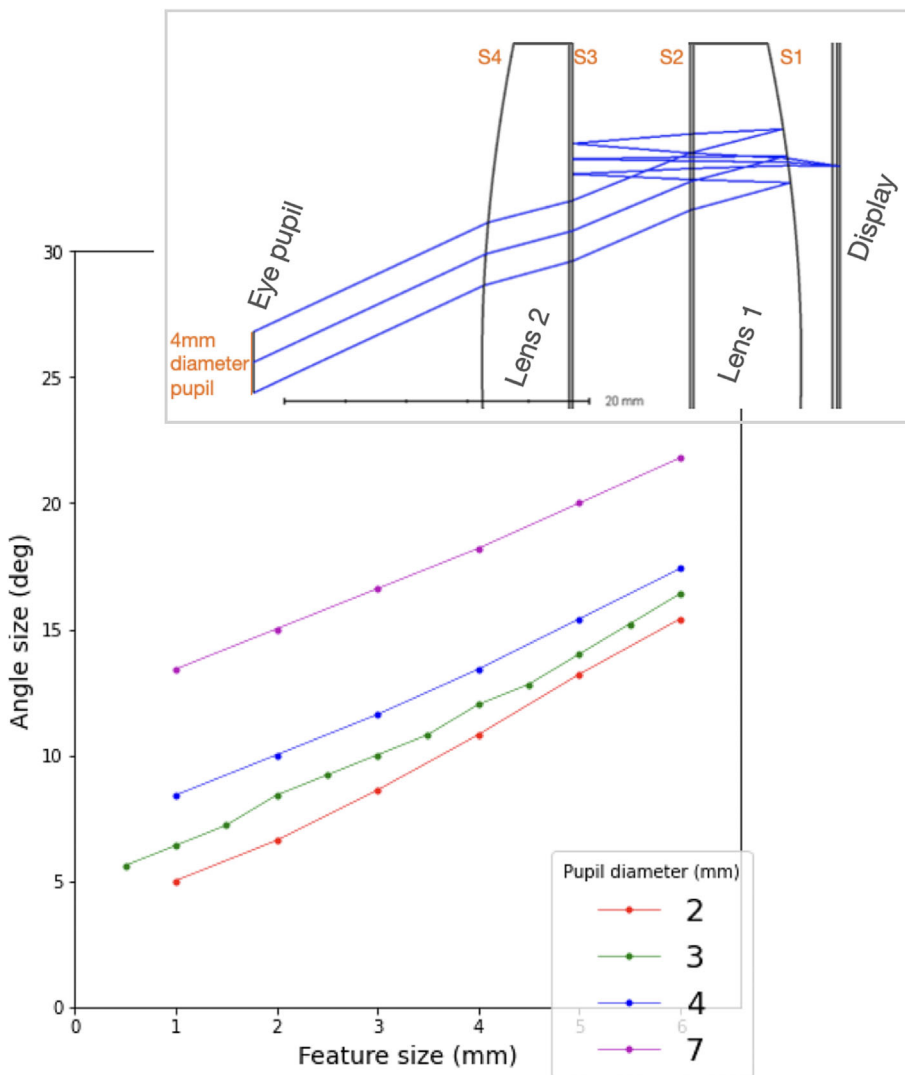

FIGURE 11   Top: an example of light beam propagating through a VR optical design. The beam size grows as it goes from display to the eye. Bottom: relation of defects' feature size to local pupil swim's angular size, at different pupil diameters. This is from simulation.
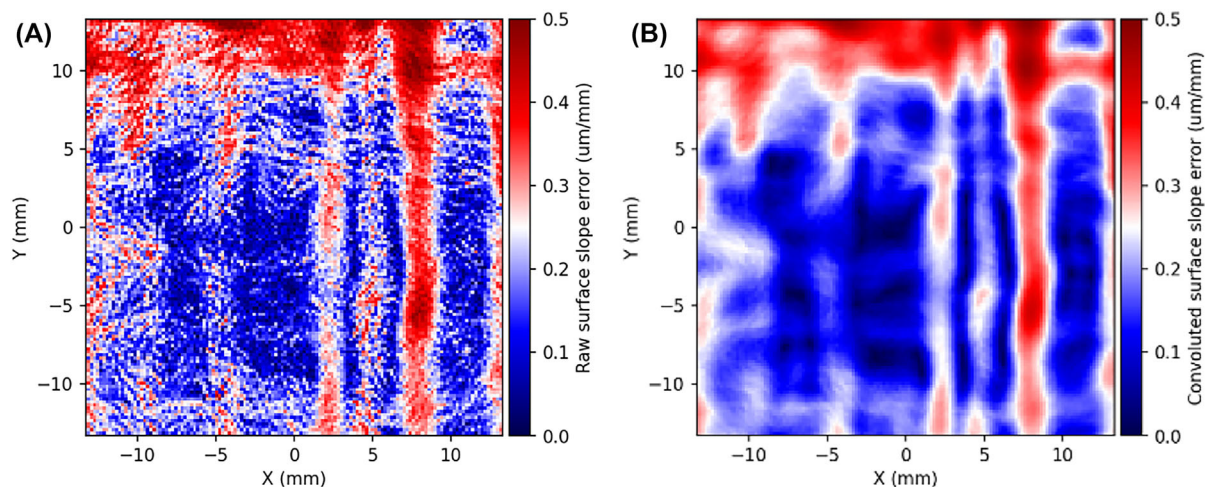
FIGURE 12 (A) Raw measured surface slope error from interferometer measurement. (B) Convoluted surface slope error with smoothing filter derived from simulation and user study data.
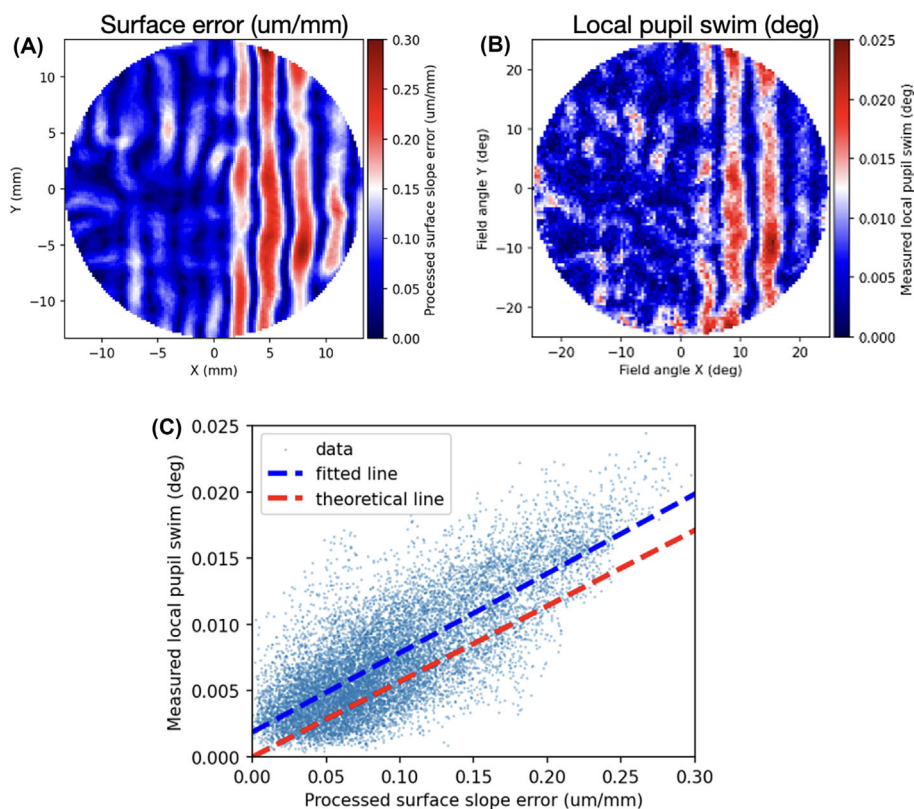

FIGURE 13 (A) Processed slope error map of a lens surface. Raw slope error was from measurement of a real sample. (B) Measured local pupil swim of the same sample. (C) Correlation between data in (A) and (B). The theoretical line is generated using Equation (3).

4.3 | Correlating surface error and local pupil swim

We evaluated several samples that have both pupil swim data and component surface error data with visible defects for our correlation study. Figure 13A showed the processed surface error slope map from one such sample, and Figure 13B showed the measured local pupil swim

from the same sample. The resemblance between Figure 13A,B across the FOV is obvious in terms of both locations and magnitudes.

From the simulation, we understood that the local pupil swim is proportional to the processed surface slope error. Given the optical system design, we further matched the experiment data with the surface coordinates (mm) and the field angles (degrees). Then, we

correlated the data and plotted in Figure 13C with surface slope error as x axis and local pupil swim as y axis. The theoretical line was predicted from Equation (3). The fitted line has only 5% difference in slope compared to the theoretical line. The Pearson's correlation coefficient was found to be 0.785 with p value $\ll 0.05$, which also showed a strong correlation. Note the measured local pupil swim in Figure 13B comprised of surface defects from four optical surfaces. We only took the most sensitive surface component data for correlation here. Since the other surfaces could lead to extra pupil swim, there can be the near-constant positive systematic error of the fitted line compared to the theoretical line.

5 | CONCLUSION

We investigated the root cause and perceptual effect of a common visual defect in VR/AR headset—local pupil swim, which appeared as unnatural ripples across virtual content. It is often generated by optical surface defects that alter virtual content direction unexpectedly. People are particularly sensitive to observe this visual effect when conducting VOR in a world-locked virtual content.

Combining the optical simulation and a user study, we have generated three metrics and unified them in a perceptual model: (1) component surface error characteristics (surface defect size, slope, and location), (2) system metric local pupil swim (with the unit of degree), and (3) user perceptual quality. The model was supported by a user study and validated by optical simulation and measurement. This model is greatly useful for product design, tolerancing, and evaluating product performance.

ORCID
Jerry Jia https://orcid.org/0000-0002-8082-2674

REFERENCES
1. McColgin FH. Movement thresholds in peripheral vision. J Opt Soc Am. 1960;50(8):774. https://doi.org/10.1364/JOSA.50.000774
2. McKee SP, Nakayama K. The detection of motion in the peripheral visual field. Vision Res. 1984;24(l):25–32.
3. Chan TT, Wang Y, So RHY, Jia J. Predicting subjective discomfort associated with lens distortion in VR headsets during vestibulo-ocular response to VR scenes. IEEE Trans vis Comput Graph. https://doi.org/10.1109/TVCG.2022.3168190
4. Melzer JE, Moffitt K. Head-mounted displays: designing for the user. 1st ed. Scotts Valley, California: CreateSpace Independent Publishing Platform; 2011.
5. Kress B. Optical architectures for augmented-, virtual- and mixed-reality headsets. Bellingham, WA: SPIE Press; 2020.
6. Lu J, Lian T, Jia J. Display system sharpness modeling and requirement in VR and AR applications. In: Electronic Imaging Conference. San Francisco, California: The Society for Imaging Science and Technology; 2023.
7. ITU-R BT. 500–11. Methodology for the subjective assessment of the quality for television pictures. 2002.
8. Fairchild M. Color appearance models. 2nd ed. Hoboken, NJ: John Wiley & Sons, Ltd.; 2005. p. 38–39.
9. Engeldrum P. Psychometric scaling: a toolkit for imaging systems development. Winchester, MA: Imcotek Press. First Printing; 2000.
10. Culler A, Elliott E. How to use the Grid Sag surface type [Internet]. [reviewed 2020 Mar 9; cited 2023 Mar 8]. Available from: https://support.zemax.com/hc/en-us/articles/1500005490241-How-to-use-the-Grid-Sag-surface-type
11. Shioiri S, Ito S, Sakurai K, Yaguchi H. Detection of relative and uniform motion. J Opt Soc Am A. 2002;19(11):2169–2179. https://doi.org/10.1364/JOSAA.19.002169
12. Franssen L, Tabernero J, Joris E, Coppens J, van den Berg T. Pupil size and retinal straylight in the normal eye. Invest Ophthalmol Vis Sci. 2007;48(5):2375–2382. https://doi.org/10.1167/iovs.06-0759
13. Spector RH. Chapter 58: The pupils. In: Walker HK, Hall WD, Hurst JW, editorsClinical methods: the history, physical, and laboratory examinations 3rd ed. Boston: Butterworths; 1990. Available from: https://www.ncbi.nlm.nih.gov/books/NBK381
14. Cheng D, Hou Q, Li Y, Zhang T, Li D, Huang Y, et al. Optical design and pupil swim analysis of a compact, large EPD and immersive VR head mounted display. Opt Express. 2022;30(5):6584–6602. https://doi.org/10.1364/OE.452747

AUTHOR BIOGRAPHIES

Jerry Jia is a system engineer and human perception specialist at Meta Reality Labs in California, USA. His role at Meta is to integrate hardware systems, algorithms, and human vision in the product development of VR/AR. His work includes mathematically modeling motion sickness, architecting mixed reality, and designing optical systems to enable delightful user features. Before Meta, he was a main

inventor and engineer at Apple for the True Tone display feature, specializing in color sensing and algorithms. He received his Ph.D. in Material Science from Columbia University in the City of New York and B.S. in Physics, B.A. in Philosophy from Peking University in Beijing.

Tsz Tai Chan received a B.Eng. in Mechanical Engineering (2017) and M.Phil. in Industrial Engineering and Decision Analytics (2021) from the Hong Kong University of Science and Technology. He has experience on various data analytics projects in manufacturing and healthcare industry. He is now working as an optical engineer at the Facebook Reality Lab. His main interests are computational models for analyzing visual data, especially 3D computer vision methods and their applications.

Trisha Lian is a research scientist at Reality Labs. Before joining Meta, she received her PhD in Electrical Engineering at Stanford University, working with Professor Brian Wandell. Her current research focuses on developing imaging system simulations alongside models of the human visual system, to better understand visual artifacts in head-mounted displays.

Kevin Rio is a research scientist on the Applied Perception Science team at Reality Labs, Meta. Previously, he was a UX researcher at Microsoft, supporting HoloLens and Windows Mixed Reality. He received his Ph.D. in Cognitive Science from Brown University and B.S. degrees in Physics and Psychology from Rensselaer Polytechnic Institute. Kevin's work leverages knowledge about human cognition and visual perception to improve the design of AR/VR devices and experiences.