

# Psychovisual Experimentation Using mLLMs As Observers

Robin Jenkin; NVIDIA; Santa Clara, USA; Preeti S Pillai, Aruna S Nayak, Abhishek A Joshi, Vasudhaika S, Sinchana C, Abhishek Patil; KLE Technological University; Huballi, India

## Abstract

*With the rapid progress of multi-modal large language models (mLLMs), there is growing interest in whether such models can act as judges of image quality. A fundamental question exists, however, as to the ability of such models to distinguish between various levels of image quality attributes, such as sharpness and noise.*

*This work represents one of the first systematic investigations of mLLMs as evaluators in classical paired comparison image quality assessment (IQA) experiments. Prior work in mLLM-based vision has focused on captioning or recognition tasks, whereas our study explicitly frames Gemini 2.0 Flash as a proxy subject in psychovisual testing to establish just noticeable differences (JNDs) for sharpness and noise using the Kodak image quality ruler dataset as stimulus.*

*For both sharpness and noise, the magnitudes of JNDs were found to be proportional to the relative quality of the stimulus. Surprisingly, judgments of individual pairs of images were found to be probabilistic rather than absolute, with more uncertainty observed for sharpness discrimination than noise. Prompt engineering is detailed as is the statistical analysis of results. Understanding the extent to which mLLMs can act as reliable perceptual proxies offers transformative implications for automated IQA, dataset labeling, and adaptive imaging pipelines.*

## Introduction

Image quality assessment (IQA) is a critical component of modern imaging and computer vision applications, including photography, medical imaging, surveillance, and image compression. Among various forms of degradation, image noise and sharpness loss are two of the most perceptually impactful distortions. Noise reduces visual clarity by introducing random variations in pixel intensities, while sharpness degradation blurs fine textures and edges. Accurate measurement of such degradations is essential for optimizing imaging systems and enhancing perceptual visual quality.

Traditionally, human subjective testing has been the standard for assessing perceptual image quality. Techniques such as the ISO 20462 softcopy quality ruler method [1] rely on human observers to identify which of the two images is preferred or of higher quality. Statistical analysis produces Just Noticeable Difference (JND) measurements, which quantify the minimal perceptual change detectable by human vision. While the methods are well-established and scientifically validated, large-scale subjective experiments require many observers, repeated trials, and controlled display conditions. Conducting human studies on large datasets is impractical. Human responses may be inconsistent due to fatigue, adaptation, or bias, requiring averaging over many participants.

To overcome these challenges, a novel approach is proposed

using multi-modal Large Language Models (mLLMs) as perceptual evaluators. The potential of mLLMs may be leveraged as a scalable alternative for perceptual evaluation through paired comparison experiments. Gemini 2.0 Flash, a multi-modal LLM capable of reasoning using multiple images, documents and text input was used to experiment with this capability. The model was prompted to evaluate image quality-based image pairs varying in noise and sharpness. The experiments are conducted using the Kodak image dataset, which is a widely recognized benchmark in perceptual image quality studies [1]. The dataset consists of diverse, high-resolution natural images with rich textures and edges, making it suitable for testing sensitivity to noise and sharpness variations. From the dataset, two experimental sets were generated: (1) images with varying levels of shot noise, and (2) images subjected to Gaussian blur at different sharpness levels. Each image variant was compared against multiple others using a structured paired comparison design. The mLLMs' output is used to construct frequency matrices, which are then converted into proportion and scale difference matrices. A Gaussian psychometric model is applied to derive JND estimates, capturing the point where changes in noise or sharpness become perceptually noticeable. The study demonstrates an AI-driven approach to perceptual image quality assessment, capable of scaling traditional psychophysical testing while capturing human-like perceptual patterns for both image noise and sharpness.

## Related Work

Many works detail the development of psychometric scaling and further its application to imaging [1], [2], [7]. Foundational work such as the ISO 20462 Quality Ruler method [1] and its softcopy implementation demonstrate the effectiveness of paired comparison techniques and usefulness of just noticeable difference (JND) scaling for subjective image quality assessment [1]. Studies have compared paired comparisons with rank ordering, highlighting their balance between accuracy, precision and consistency in human assessment tasks. Additionally, system-level insights on environment setup, observer consistency, and interface design provide valuable guidance for applying similar psychophysical methods in new domains. This work builds upon these foundations, extending paired comparison methodologies to the evaluation of mLLM-generated observations.

In an effort to enhance both usability and scalability, Jin et al. [1] developed a digital version of the quality ruler method, substituting physical prints with reference images displayed on a screen. This system allows users to compare a test image against a series of calibrated reference images that depict different levels of perceived sharpness, as measured by the Standard Quality Scale (SQS) [1]. Through a graphical user interface (GUI), users can navigate through the reference levels and choose the one that best matches the quality of the test image. This digital approach not

only showed high consistency among different observers but also cut evaluation time by about 50% compared to traditional hard-copy methods.

In this work, Jin and Keelan [1] formalized a lot of the necessary system requirements for the effective implementation of IQA using displays. These requirements include high-resolution display specifications, calibrated lighting conditions, controlled viewing distances with the use of headrests, and standardized room configurations. They also discussed software components, such as MATLAB-based tools for presenting stimuli and logging outputs, which are crucial for ensuring consistency across evaluation sessions. One of the critical outputs of the work was the standard reference stimulus, evaluated by experts, to yield a set of images calibrated to JNDs of image quality loss for sharpness. By measuring image quality attributes against these reference stimuli, experimentation can be compared against a standard ruler of image quality loss, hence their choice in this study.

Beyond system implementation, comparative studies of psychophysical evaluation methods have shed light on the trade-offs between different methodologies. For example, Cui [7] examined the paired comparison and rank ordering techniques for assessing colour image quality. Although both methods yielded similar rank orderings, paired comparisons were found to be more accurate, albeit at a slower pace. The research pointed out that while rank ordering is efficient, it can increase cognitive load and decrease consistency in judgments, making paired comparisons more appropriate for applications that demand high precision.

Jin et al. further validated the softcopy ruler method in a multi-laboratory environment [2], confirming the consistency of SQS ratings among different observers and institutions. Burns et al. [5] subsequently adapted the ISO 20462 method for online image quality studies, demonstrating the feasibility of conducting quality ruler assessments in less controlled environments. The results of these studies underscored the reliability of the softcopy ruler as a psychophysical tool, providing an effective framework for quantifying perceptual quality through human judgment. To ensure the reliability of observer responses, basic vision screening procedures are often advised. Schneider [9] offered a printable eye chart for assessing visual acuity, which has been used in several subjective evaluation studies to ensure participants can discern image details at specified viewing distances.

The emergence of deep learning and foundation models has prompted investigation of computational approaches to IQA. Cheon et al. [4] demonstrated the use of transformer architectures for perceptual image quality assessment, while Pérez-Ortiz et al. [6] developed methods for constructing unified quality scales from pairwise comparisons and ratings. More recently, the rapid advancement of mLLMs has opened new possibilities for automated quality evaluation. Cui et al. [3] proposed M3-AGIQA for multimodal assessment of AI-generated image quality, and Wu et al. [10] conducted a comprehensive study of mLLM prompting strategies for IQA using psychophysics testing procedures, finding that models such as GPT-4V provide reasonable correlation with human perception but struggle with fine-grained quality discrimination. Despite this growing body of work, no prior study has applied mLLMs as proxy observers within the classical paired comparison framework that underpins standardized psychovisual testing such as ISO 20462. Building on these foundations, the current study adapts the paired comparison methodology and JND-

based scaling principles to evaluate observations by mLLMs, aiming to take the first small steps toward a robust and interpretable framework for evaluating images using mLLMs that correspond to human preference.

## Methodology

The work introduces a prompt-based framework that utilizes an mLLM via an API to evaluate image quality attributes of noise and sharpness using a paired comparison methodology. The high-level experimental workflow is shown in Figure 1, and begins by selecting a pair of images from a prepared set which varies in severity of the image quality attribute to be measured. For these first experiments sharpness and noise were varied independently to generate two sets of images.

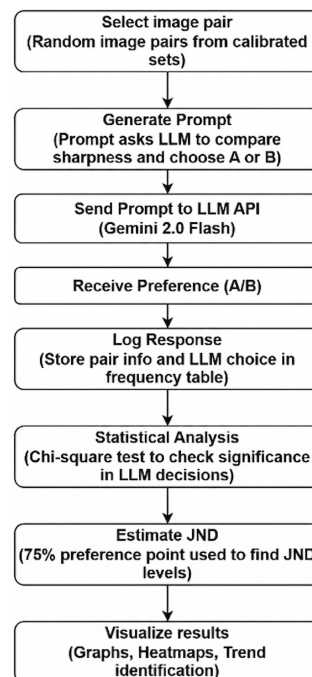


Figure 1. Experimental Workflow

For noise analysis, 12 images from the SQS ruler reference were degraded using known levels of shot noise, yielding SNR of 3 to 36dB in 3dB steps. For the sharpness attribute, 8 images from the ruler reference were chosen, each having previously been degraded into 16 calibrated levels of sharpness, representing 0 to 30 JNDs of quality as defined by ISO20462[1]. Figures 2 and 3 show example images and treatments used in the study.

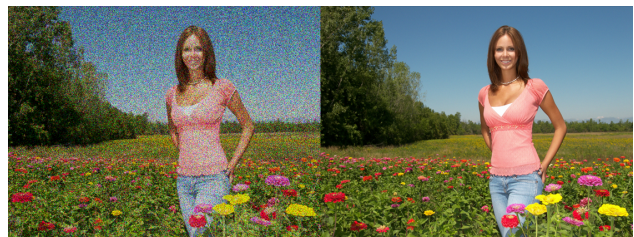


Figure 2. Example of two noise treatments with SNR of 3dB and 36dB respectively.



**Figure 3.** Example of two sharpness treatments corresponding to 15 and 0 JNDs of quality loss respectively, as defined by ISO20462[1].

The study employed Gemini 2.0 Flash, an mLLM, with a random image pair generator to select image pairs across the datasets. The selected pair is tested with a carefully crafted prompt that insisted that the mLLM compare the two images (Image 1 and Image 2) based on noise or sharpness, explicitly ignoring image content and including cases with minimal or subtle differences.

Preliminary experiments revealed that the model occasionally provided ambiguous responses, such as "neither", which compromised the evaluation process. To address this limitation, the prompt was iteratively refined to enforce a binary decision between the two images. It was also found that the mLLM would bias answers to one image when it could not distinguish a difference. As for human observers this reinforced the need to randomly assign images to either image 1 or image 2 when presented. Further, filenames of images were changed to remove any reference or ordering to mask the attribute strength completely from the mLLM. The final prompt formulation compelled the model to make a definitive choice even when faced with uncertainty, effectively eliminating non-committal responses and ensuring consistent comparative assessments across all image pairs. The final prompt arrived at was:

You are given two images. Ignore their filenames and content entirely. Evaluate each image entirely based on the quality factor that is Noise level. Select the image that has the superior quality among the two. Return only 'Image 1' or 'Image 2' as your response. Even if the images appear indistinguishable, you must return either 'Image 1' or 'Image 2' by making a guess.

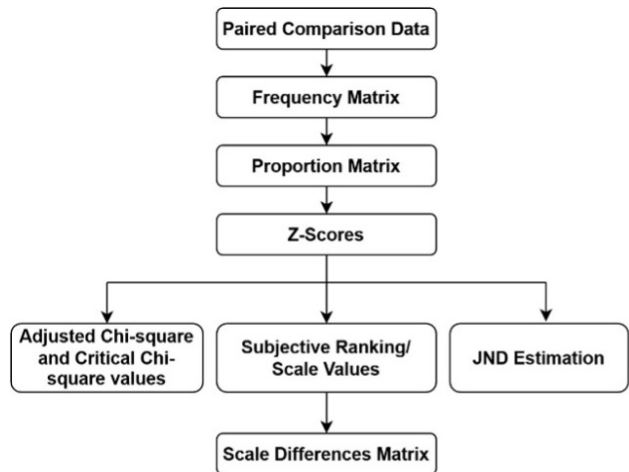
Noise level is substituted by sharpness level for the sharpness based experimentation.

The combined prompt and image pair is sent to the mLLM via an API call. Each discrete level of an image was compared 12 times with one another yielding approximately 21,000 comparisons for noise analysis and 13,500 comparisons for sharpness analysis. This highlights one of the advantages of performing image comparison using mLLMs, extensive studies may be undertaken as they do not fatigue. The noise experimentation would take approximately 58 hours of comparison at just 10 seconds per pair, with no breaks, if a human were to participate. The model's preferences are logged in a frequency table, where each cell (i,j) represents the number of times image j was preferred over image i.

The statistical analysis flow is illustrated in Figure 4 and

based on Engeldrum[8]. The analysis assesses whether the mLLMs' preference is distinguishable from random behavior, thus determining the presence of perceptual preference. The raw paired comparison data captured in the frequency matrix is converted into a proportion matrix, which indicates the probability of one image being preferred over the other. This is calculated simply by dividing the values in column (i,j) by the total number of times the pair (i,j|i) was compared. Figures 5 and 6 detail the frequency matrices compiled for all images for the noise and sharpness experiments respectively.

For conversion of the proportion matrix to z-scores, Thurstone Case V is assumed [8], such that the sensation produced by each stimulus is normally distributed. The difference in sensation (stimulus i minus stimulus j) is then also normal. So the proportion P(i,j) corresponds to a z-score:  $Z(i,j) = \text{norminv}(P(i,j))$ . We clip P away from 0 and 1 to avoid infinite Z. Z(i,j) is then the standardized difference in sensation. Subjective rankings, scale differences and JND estimation are then derived from the z-scores. A Chi-Square test was conducted to yield adjusted and critical chi-square values to evaluate the statistical significance of the results.



**Figure 4.** The statistical analysis flow based on Engeldrum [8].

Treatment Levels dB	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	36
3	0	94	125	136	154	156	156	156	156	156	156	156	156	156	156	156	156
5	62	0	98	127	146	156	156	156	156	156	156	156	156	156	156	156	156
7	31	58	0	93	105	127	156	156	156	156	156	156	156	156	156	156	156
9	20	29	63	0	106	107	138	146	155	156	156	156	156	156	156	156	156
11	2	10	51	49	0	90	111	141	156	156	156	156	156	156	156	156	156
13	0	0	29	49	62	0	96	132	154	156	156	156	156	156	156	156	156
15	0	0	0	17	44	58	0	116	136	148	156	156	156	156	156	156	156
17	0	0	0	6	15	24	40	0	107	147	156	154	156	156	156	156	156
19	0	0	0	1	0	2	20	50	0	118	148	146	152	155	153	154	156
21	0	0	0	0	0	0	8	9	38	0	116	132	141	140	137	142	150
23	0	0	0	0	0	0	0	0	8	40	0	97	110	114	122	136	132
25	0	0	0	0	0	0	0	2	10	24	59	0	106	116	121	125	125
27	0	0	0	0	0	0	0	0	4	15	46	50	0	88	84	104	106
29	0	0	0	0	0	0	0	0	1	16	42	40	68	0	84	93	102
31	0	0	0	0	0	0	0	0	3	19	34	35	72	72	0	78	82
33	0	0	0	0	0	0	0	0	2	14	20	31	52	63	78	0	83
36	0	0	0	0	0	0	0	0	0	6	24	31	50	46	74	73	0

**Figure 5.** The frequency matrix for all images for the noise treatments. Treatments are specified in SNR (dB).

Scale values for the  $j^{th}$  treatment may be determined by cal-

culating the mean of z-scores in column j.

$$\text{scale}(j) = \frac{1}{N} \sum_{i=1}^N z(i, j) \quad (1)$$

where N is the number of samples and z(i,j) corresponds to the case where sample j is preferred over sample i. Each entry in the scale difference matrix reflects the perceptual gap between two images, quantifying not only the direction of preference but also the magnitude of the difference. The mean of all z-scores is typically subtracted from the results to center the values around zero.

Scale differences are calculated simply as the difference between the scale values for images i and j for each combination, scale difference(i, j) = scale(j) - scale(i).

A JND represents the minimum change in stimulus intensity that an observer can reliably perceive. In the paired comparison experiments, a JND is quantified by modeling the relationship between stimulus strength and the probability of detection using a psychometric function. Using a Gaussian psychometric function, the response probability P(x) at stimulus level x is given as,

$$P(x) = \Phi(\alpha + \beta x) \quad (2)$$

where,  $\Phi$  is the standard normal cumulative distribution function, while  $\alpha$  and  $\beta$  represent model parameters estimated from the mLLMs' responses.

A JND calculated where an observer is correct 75% of the time corresponds to a z-distance of 0.674 [8]. Using the calculated z-scores we estimate the distance in treatment units of the stimulus that correspond to a change in z-score of 0.674. A smaller JND indicates that an observer is highly sensitive to differences in image quality, while a larger JND indicates reduced sensitivity.

Though nascent, this approach facilitates the estimation of perceptual sensitivity across different image quality attributes for mLLMs as observers. Further comparison to the performance of humans will enable their use as proxies in IQA tasks providing a reproducible and scalable alternative.

## Results and Discussion

Figures 5 and 6 are the frequency matrices for all images for both noise and sharpness treatments. A cursory glance at these

Treatment Levels	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30
0	0	63	79	75	77	72	72	62	82	72	79	82	72	71	77	79
2	50	0	78	76	79	74	77	73	70	65	77	68	75	82	80	67
4	34	35	0	66	77	71	80	70	75	73	73	79	73	66	82	74
6	37	38	46	0	73	70	72	83	69	71	75	72	82	58	75	84
8	35	33	35	38	0	72	73	78	81	81	69	68	64	76	62	73
10	40	38	41	42	42	0	71	72	68	80	77	80	69	65	66	74
12	40	35	33	41	39	39	0	74	84	75	77	69	73	62	65	75
14	50	39	44	30	34	38	39	0	74	82	76	82	73	75	66	70
16	30	43	37	44	31	45	28	38	0	79	80	77	76	76	72	69
18	40	47	40	41	32	33	37	30	33	0	65	71	67	77	78	75
20	33	35	40	38	44	36	35	36	31	43	0	76	79	72	75	75
22	32	44	33	40	46	32	43	30	35	39	38	0	77	81	73	75
24	40	37	37	30	48	44	39	39	36	43	31	34	0	71	74	70
26	42	31	47	54	37	48	50	37	36	35	38	30	41	0	69	70
28	36	33	31	38	50	47	47	47	40	34	36	39	36	42	0	76
30	34	45	38	28	39	38	37	43	43	37	37	37	43	42	37	0

Figure 6. The frequency matrix for all images for the sharpness treatments. Treatments are specified in JNDs of softcopy ruler quality.

indicates that the mLLM has an easier time distinguishing noise treatments than sharpness due to the increased separation between the values in the top right of the matrix when compared to the bottom left. The closer values indicate that the mLLM finds it difficult to choose either image i or j as the better one. The diagonal remains zero, as no image was compared with itself, though sometimes this is included as a good diagnostic test of bias in the presentation of stimuli.

Figure 7 shows a graph of the z-score derived for the noise comparisons of all images versus the stimulus strength in dB. Also shown is a linear fit with an  $R^2 = 0.9595$ . The gradient of the fit implies a constant JND = 5.006 dB. Figure 8 documents JNDs calculated from the z-score graph at different stimulus strengths. The strong linear response between 3 and 25 dB SNR confirms a Weber-Law-like proportionality in this regime, where a consistent change in stimulus strength yields a consistent JND for the system. Between approximately 15 and 20 dB the mLLM becomes marginally better at distinguishing changing noise, with peak sensitivity dropping the JND to approximately 3 dB. Above image SNR of about 25dB the mLLM finds it increasingly difficult to 'see' differences in noise between the images with a JND rising to approximately 25dB for images with SNR of 30 to 36dB. The scale S flattens and saturates above approximately 25 dB SNR. This saturation is likely attributable to the internal tokenization architecture of the mLLM. Gemini's documentation notes that "image inputs with both dimensions  $\leq 384$  pixels are counted as 258 tokens. Images larger in one or both dimensions are cropped and scaled as needed into tiles of  $768 \times 768$  pixels, each counted as 258 tokens." This downsampling inherently discards the fine pixel-level detail needed to distinguish subtle noise differences at high image quality.

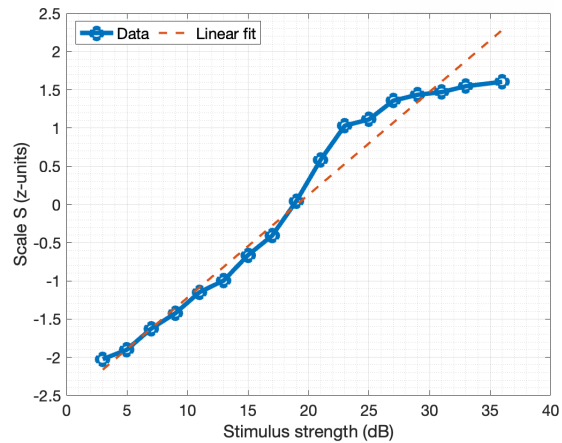


Figure 7. Z-scores versus stimulus strength for all images for noise as observed by Gemini 2.0.

Figures 9 and 10 show the z-scores and JNDs for each individual image. While there appears to be some sensitivity to image content as evidenced by the variation in the curves, the majority of curves exhibit this increase in JND and thus a decreased ability in the mLLM to distinguish quality changes due to noise at these high quality levels. The sharp JND rise above 25 dB is robust across all content types, confirming that the saturation effect is attributable to the mLLM architecture rather than image-specific

factors.

Figure 11 shows a graph of the z-score derived for the sharpness comparisons of all images versus the stimulus strength in quality ruler JNDs. Also shown is a linear fit with an  $R^2 = 0.9934$ . The gradient of the fit implies a constant JND = 25.00 JND (SQS). Figure 12 documents JNDs calculated from the z-score graph at different stimulus strengths. While the measured z-scores are a good fit to the linear model, it may be seen that the mLLM is a poor judge of sharpness as anticipated by the frequency matrix. The 75% JND varies between 20 and 30 JND(SQS) units. Given that the stimulus treatments are calibrated in JNDs of quality loss as observed by humans for the development of ISO20462[1], this indicates that mLLM sensitivity is approximately  $25\times$  lower than human observers at judging the differences in sharpness. As with the noise results, the tokenization and internal downsampling of the mLLM likely limits its high-frequency perception, which is precisely the information needed to judge sharpness. It should also be noted however, that because of the rapid development of mLLMs this performance is likely to change rapidly with the release of newer models such as Gemini 3.5.

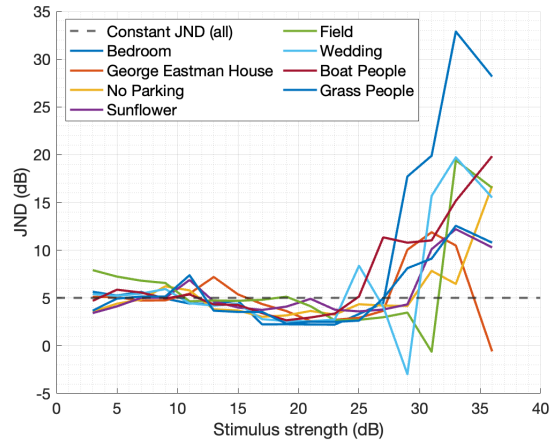


Figure 10. Calculated JNDs versus stimulus strength for individual images for noise as observed by Gemini 2.0.

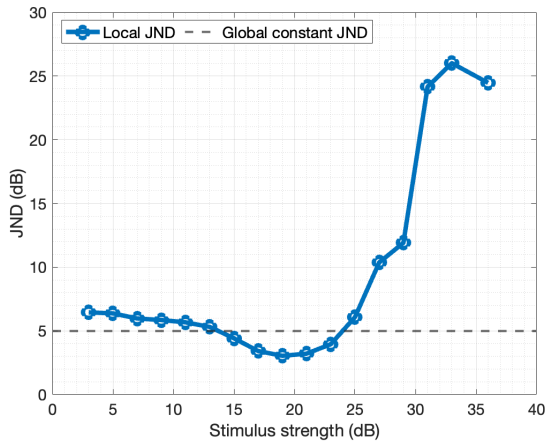


Figure 8. Calculated JNDs versus stimulus strength for all images for noise as observed by Gemini 2.0.

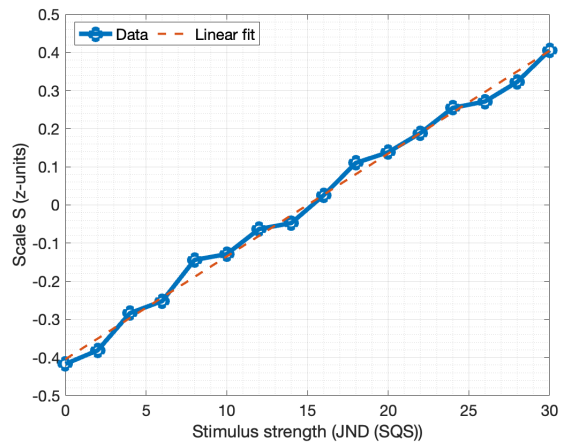


Figure 11. Z-scores versus stimulus strength for all images for sharpness as observed by Gemini 2.0.

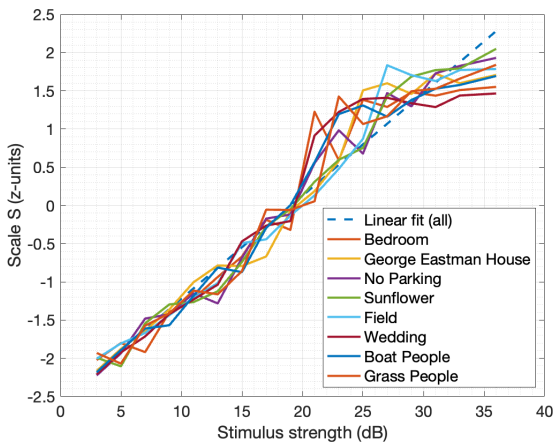


Figure 9. Z-scores versus stimulus strength for individual images for noise as observed by Gemini 2.0.

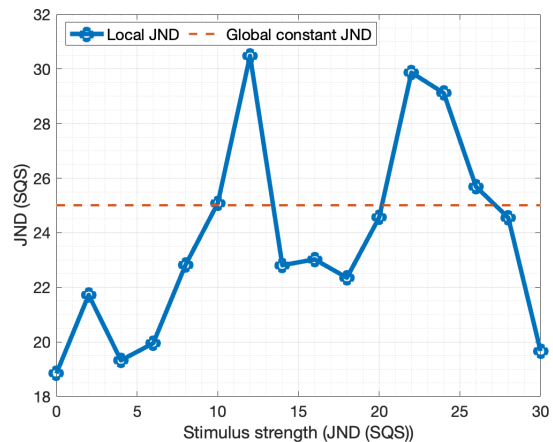
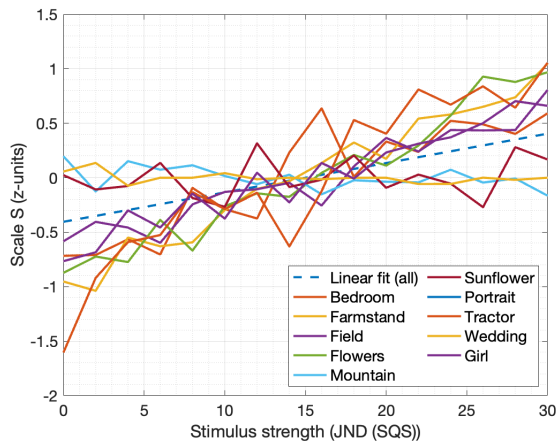


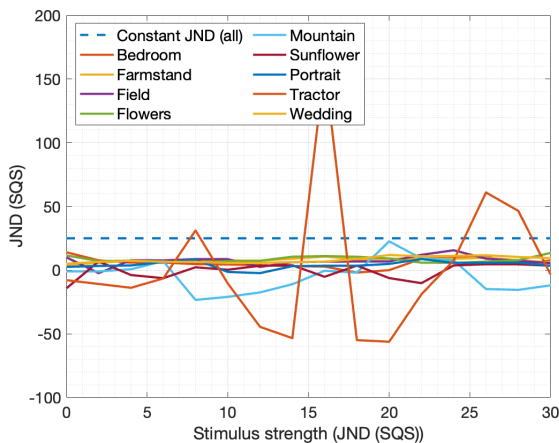
Figure 12. Calculated JNDs versus stimulus strength for all images for sharpness as observed by Gemini 2.0.

Figures 13 and 14 show the z-scores and JNDs for each individual image for the sharpness results. It may be seen that there are two distinct groupings of images in Figures 13 with different gradients. This may warrant future investigation as it demonstrates significant sensitivity to the image content. Figure 14 further shows the sensitivity of the mLLM to the 'Wedding' image with extreme variance in the measured JNDs across the range of stimulus input.



**Figure 13.** Z-scores versus stimulus strength for individual images for sharpness as observed by Gemini 2.0.

The chi-square calculated for all results for the noise study was 828.17 against a critical value of 146.57. For the sharpness study 549.23 against a critical value of 129.92.



**Figure 14.** Calculated JNDs versus stimulus strength for individual images for sharpness as observed by Gemini 2.0.

## Conclusion

This work demonstrates that mLLMs can autonomously complete psychovisual paired comparison experiments, producing statistically significant preferences for both noise and sharpness attributes. Gemini 2.0 Flash completed 34,500 comparisons, equivalent to over 58 hours of sustained human testing, at no fatigue cost.

For noise, the mLLM exhibited robust, Weber-Law-like discrimination with a global constant JND of approximately 5 dB across the active SNR range. Notably, responses were probabilistic rather than deterministic, with variability that mirrors human observer behaviour. The chi-square analysis confirmed the statistical significance of the results for both noise ( $\chi^2 = 828.17$ , critical value 146.57,  $\alpha < 0.001$ ) and sharpness ( $\chi^2 = 549.23$ , critical value 129.92,  $\alpha < 0.001$ ).

For sharpness, discrimination was poor relative to human observers, with the mLLM exhibiting approximately  $25\times$  higher JNDs. The internal tokenization architecture of Gemini, which downsamples images to tiles of  $768\times 768$  pixels, likely limits the high-frequency perception needed for sharpness judgments.

Future work should test newer models such as Gemini 3.1, GPT-4o, and Claude Opus 4.6 as they become available, and explore additional quality attributes including compression artefacts and colour accuracy. Integration of mLLM-based evaluation into automated IQA labeling pipelines for foundation model training represents a promising direction for scalable perceptual quality assessment.

## Acknowledgments

The authors would like to thank Balaji Holur, NVIDIA, who established the KLE mentorship program under which this work was undertaken.

## References

- [1] Elaine W. Jin, Brian W. Keelan, Slider-Adjusted Softcopy Ruler for Calibrated Image Quality Assessment, *Journal of Electronic Imaging*, Vol. 19, No. 1, 2010, pp. 011009-1–011009-12.
- [2] Elaine W. Jin, Brian W. Keelan, Junqing Chen, Jonathan B. Phillips, Ying Chen, Softcopy Quality Ruler Method: Implementation and Validation, *Proc. SPIE 7242*, 2009, pp. 1–14.
- [3] Chuan Cui, Kejiang Chen, Zhihua Wei, Wen Shen, Weiming Zhang, Nenghai Yu, M3-AGIQA: Multimodal, Multi-Round, Multi-Aspect AI-Generated Image Quality Assessment, *arXiv:2502.15167*, 2025, pp. 1–24.
- [4] Manri Cheon, Sung-Jun Yoon, Byungyeon Kang, Junwoo Lee, Perceptual Image Quality Assessment with Transformers, *Proc. IEEE/CVF CVPR Workshops*, 2021, pp. 433–442.
- [5] Peter D. Burns, Jonathan B. Phillips, Don Williams, Adapting ISO 20462 Softcopy Quality Ruler Method for On-Line Image Quality Studies, *Proc. SPIE 8653*, 2013, pp. 1–10.
- [6] María Pérez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, Rafal K. Mantiuk, From Pairwise Comparisons and Rating to a Unified Quality Scale, *IEEE Transactions on Image Processing*, 2019, pp. 1–14.
- [7] Chengwu Cui, Comparison of Two Psychophysical Methods for Image Color Quality Measurement: Paired Comparison and Rank Order, *IST/SID Eighth Color Imaging Conference*, 2000, pp. 222–227.
- [8] Peter G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Imcotek Press, Winchester, MA, 2000.
- [9] Joel Schneider, *Printable Eye Chart*, 2002. Available: <https://seeing.org/downloads/tools/snellen-eye-chart.pdf>
- [10] Tianhe Wu, Kede Ma, Jie Liang, Yujiu Yang, Lei Zhang, A Comprehensive Study of Multimodal Large Language Models for Image Quality Assessment, *Proc. European Conference on Computer Vision (ECCV)*, 2024, pp. 1–16.

## Author Biography

*Robin Jenkin received, BSc(Hons) Photographic and Electronic Imaging Science (1995) and his PhD (2001) in the field of image science from University of Westminster. He also holds a M.Res Computer Vision and Image Processing from University College London (1996). Robin is a Fellow of The Royal Photographic Society, UK, and Executive Vice President of Society for Imaging Science and Technology. Robin is secretary of the IEEE P2020 Image Quality for Autonomous Vehicles working group. At NVIDIA Robin is a distinguished engineer and models image quality for autonomous vehicle and other applications. He is a Visiting Professor at University of Westminster within the Computer Vision and Imaging Technology Research Group and co-author of the 10th ed. "The Manual of Photography", Focal Press.*

*Preeti S. Pillai is an Assistant Professor in the Department of Electronics and Communication Engineering at K.L.E. Technological University, India. She holds an M.Tech in VLSI Design and Embedded Systems and is currently pursuing her Ph.D., demonstrating her strong commitment to research and academic excellence. With over 13 years of teaching experience, her academic work spans embedded systems, VLSI design, and emerging electronic technologies. She has contributed to scholarly research through publications in peer-reviewed conferences and journals, including work on automotive technologies and innovative approaches to embedded systems learning. Her professional interests include engineering education, applied electronics, and fostering practical learning environments that bridge theory and real-world applications. She remains actively engaged in academia through teaching, research, and mentorship of aspiring engineers.*

*Aruna S. Nayak is an Associate Professor in the Department of Computer Science and Engineering at K.L.E. Technological University, India. She holds an M.Tech in Digital Electronics and brings over three decades of teaching experience to the field of engineering education. Her academic expertise spans embedded systems, computer organization and architecture, microcontrollers, analog and digital circuits, and the Internet of Things (IoT). She has contributed to the scholarly community through publications in peer-reviewed journals and international conferences. Her professional interests center on engineering education, applied electronics, and the design of experiential learning environments that effectively connect theoretical concepts with real-world engineering practice. She continues to be actively engaged in academia through teaching, research, and guiding young minds.*

*Abhishek A Joshi received his Bachelor's degree in Electronics and Communication Engineering from KLE Technological University. He has authored a research paper published with Springer on oral disease detection using multimodal fusion techniques.*

*Vasudhaika S received her Bachelor's degree in Electronics and Communication Engineering from KLE Technological University. She has authored a research paper published with Springer on oral disease detection using multimodal fusion techniques.*

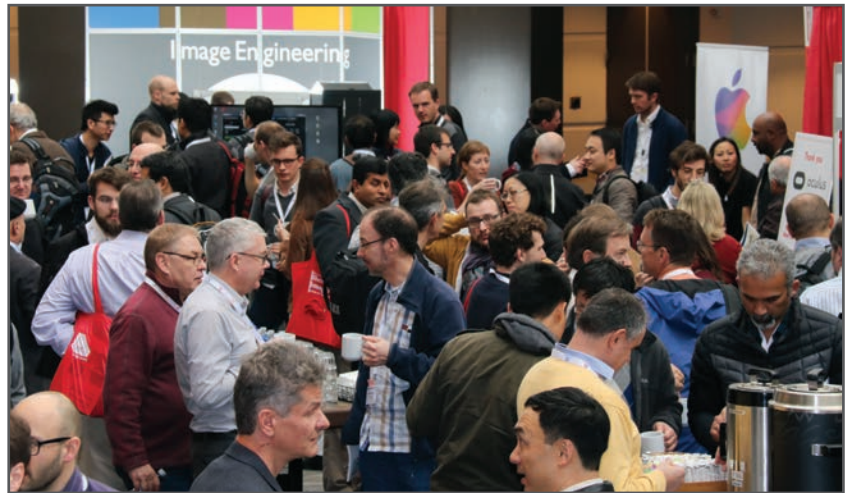
*Sinchana C received her Bachelor's degree in Electronics and Communication Engineering from KLE Technological University. She has authored a research paper published with Springer on oral disease detection using multimodal fusion techniques.*

*Abhishek Patil received his Bachelor's degree in Electronics and Communication Engineering from KLE Technological University.*

**JOIN US AT THE NEXT EI!**

# electronic IMAGING

*Imaging across applications . . . Where industry and academia meet!*



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

[www.electronicimaging.org](http://www.electronicimaging.org)

