Objective QoE Models for Streamed Social Extended Reality Applications over Heterogeneous Access Networks

Omid Saedi, Henrique Souza Rossi, Karan Mitra
Pervasive and Mobile Computing
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology, Skellefteå, Sweden
omid.saedi@ltu.se

***Abstract*—Social extended reality (Social XR) enables users to interact via embodied avatars in online shared environments, thereby offering immersion and co-presence for remote collaboration. Streamed social XR applications are highly sensitive to network impairments that may degrade end users' quality of experience (QoE). Although objective QoE models exist for conventional multimedia services, no validated models have been established for social XR. Therefore, QoE prediction via measurable quality of service (QoS) metrics remains a challenge. In this paper, we develop and validate objective QoE models (linear, polynomial, and non-linear) for social XR based on a dataset from 28 users, considering 20 network conditions. All three models satisfy the residual-normality test and generalize under leave-one-out cross-validation; the polynomial model provides the best fit ($R^2 = 0.94$, MAE $= 0.12$ MOS), while the nonlinear model achieves comparable fit with fewer parameters. A Shapley analysis identifies jitter as the dominant factor for QoE-aware provisioning in access networks. Applied to seven days of measurements over wired, 4G, and 5G Standalone networks, the polynomial model estimates QoE directly from routinely measured QoS, yielding predicted MOS of 3.4–4.2 for wired, 2.6–3.0 for 5G Standalone, and 2.2–2.3 for 4G networks. These results indicate shortcomings of existing mobile networks for streaming social XR applications.**

***Index Terms*—Extended Reality, Quality of Experience, Quality of Service, Subjective Tests, Prediction, Virtual Reality**

I. Introduction

Quality of experience (QoE) is a metric that represents the overall satisfaction regarding an application or a service as perceived by end users, and has become the principal criterion by which networked multimedia services are evaluated [1]. For stakeholders such as network operators and cloud service providers, subjective QoE measurement is inherently challenging. For example, obtaining valid user ratings requires extensive subjective testing, which is costly given the large number of tests needed to cover the full range of heterogeneous access-network (HANs) conditions. What these stakeholders can observe and provision instead is quality of service (QoS): objective network-level parameters such as delay, jitter, and packet loss that are usually reliably accessible. Objective QoE models reconcile this disparity by inferring perceived quality from measurable QoS parameters, and thereby support QoE-aware network management. Such models are particularly relevant for services with low tolerance to network impairment.

Streamed social extended reality (Social XR) is one such case where immersive multi-user applications are rendered in the cloud and delivered to thin clients such as head-mounted displays over HANs such as Wi-Fi, 4G, and 5G [2]–[5]. Because the interaction in social XR is real-time and the traffic is bidirectional, degradation on either the uplink or downlink can affect the QoE. As an enabling tool for collaboration and socializing [6], [7], such applications impose stringent demands on the underlying network. Recent subjective studies have begun to characterize how network conditions affect QoE in this setting [5], [8], [9]. Beyond such characterization, however, a predictive mapping from network QoS to QoE is needed for tasks such as rendering-server placement, and quality-aware resource allocation. The studies reviewed in Section V highlight the lack of validated objective QoE models for streamed social XR based on measurable network QoS. Developing such an objective QoE model is not straightforward. The relationship between QoS and QoE can be nonlinear and shaped by interactions between QoS factors such as delay and jitter. The predicted quantity is a mean opinion score (MOS) usually confined to the interval $[1, 5]$ which an unconstrained model may exceed. Subjective datasets are, moreover, small, so careful model selection and validation are essential to avoid overfitting. We draw on the subjective dataset of [5][1], in which 28 users rated a cloud-streamed physiotherapy application under 20 controlled network conditions spanning round-trip time (RTT), packet loss (PL), and random jitter (RJ).

Contributions: The paper makes three contributions: (i) it develops and validates linear, polynomial, and nonlinear objective QoE models for streamed social XR over heterogeneous access networks; (ii) it uses Shapley values to quantify the contribution of RTT, PL, and RJ to each model's predictions; and (iii) it applies the selected model to week-long wired, 4G, and 5G Standalone measurements to estimate social XR QoE under measured access-network conditions.

[1]This work was conducted as part of the development of the ITU-T P.1321 Recommendation [10].

TABLE I
NETWORK CONDITIONS IN THE DATASET [5].

Factor	# Cond.	Values
RTT (ms)	5	4, 27, 52, 77, 152
RTT (ms), PL (%)	6	[27; 2,6], [52; 2,6], [77; 2,6]
RTT (ms), RJ (ms)	9	[27; 1,3,6], [52; 1,3,6], [77; 1,3,6]
Total: 20 (listed RTT values include the testbed's 2 ms baseline)		

II. DATASET

Our models are developed using the datasets gathered by Mitra et al. [5]. Their study involved 28 participants in a two-person cloud-streamed physiotherapy session. In each session, a participant (subject) mirrored the therapist's movements while both of them maintained continuous visual and verbal communication from separate rooms. The Resonite application was rendered on a server and delivered through SteamVR to a Meta Quest 3 headset, while a laboratory testbed imposed controlled conditions with NetEm, applied symmetrically to the up- and down-links. After each condition, participants rated their QoE on a five-point absolute category ratings (ACR) scale following recommended practice for interactive VR [11]; once outliers were removed, a per-condition MOS was obtained by averaging ratings. The 20 conditions, listed in Table I, are organized as three sweeps: RTT in isolation (five levels, 4–152 ms), RTT paired with packet loss (six combinations), and RTT paired with jitter (nine combinations), together spanning the operating envelope of current fixed and mobile broadband. These 20 MOS values constitute the entire training set for the models below.

III. QOE MODELS FOR SOCIAL XR

The objective is to estimate QoE measured via MOS (MOS_{QoE}) from the measurable network parameters. Each observation is represented by a set of $(\text{RTT}, \text{PL}, \text{RJ})$, where RTT is measured in milliseconds, PL in percent, and RJ as standard deviation of RTT measured in milliseconds (ms). The response variable is the per-condition MOS obtained from the subjective dataset described in Section II. We developed three objective models: linear, polynomial and non-linear. The linear and polynomial models were fitted by least squares, using terms derived from RTT, PL, and RJ. Candidate polynomial specifications included main effects, higher-order terms and pairwise interactions; terms were then pruned according to residual behavior and coefficient significance. The non-linear model was fitted using non-linear least squares and follows an impairment-factor structure in which network degradations reduce a maximum attainable MOS. The three retained models therefore represent increasing modeling assumptions: an interpretable linear baseline, a polynomial model that can capture curvature and interactions, and a bounded impairment model whose output remains within the MOS range. Their fitted coefficients are reported in Table II.

TABLE II
MODELS' FITTED COEFFICIENTS.

Model	Coefficients
Lin.Reg.	l_0=4.2252, l_1=−0.010940, l_2=0.057596, l_3=−0.222871, l_4=−0.0025755
Poly.Reg.	p_0=4.6094, p_1=−0.024468, p_2=−0.304049, p_3=−0.0012487, p_4=0.0018482, p_5=8.03×10^{-5}
NonLin.Reg.	Q_{Max}=4.4839 (fixed), n_1=0.0055178, n_2=0.0271769, n_3=0.1097888

Linear (Lin.Reg.): Equation (1) regresses MOS on the three main effects and an RTT·PL interaction. RTT·PL term lets the effect of packet loss depend on latency, since packet loss and delay compound each other's impact on streamed interactive rendering.

$$x = l_0 + l_1 \cdot \text{RTT} + l_2 \cdot \text{PL} + l_3 \cdot \text{RJ} + l_4 \cdot (\text{RTT} \cdot \text{PL}) \quad (1)$$

Polynomial (Poly.Reg.): The polynomial specification was selected from candidate quadratic and cubic terms and their interactions by retaining terms that reduced residual error while keeping statistically significant coefficients. Equation (2) adds an RTT^2 term, fitted over the full RTT sweep, and RTT·RJ and RTT·PL interactions (RTT $\leq$ 77 ms, with RJ $\leq$ 6 ms and PL $\leq 6\%$). This form is used to represent the curvature observed in the RTT sweep and the interaction between delay and jitter in the combined RTT–RJ conditions, while avoiding higher-order terms unsupported by the sample size.

$$\begin{aligned} x = p_0 + p_1 \cdot \text{RTT} + p_2 \cdot \text{RJ} + p_3 \cdot (\text{RTT} \cdot \text{PL}) \\ + p_4 \cdot (\text{RTT} \cdot \text{RJ}) + p_5 \cdot \text{RTT}^2 \end{aligned} \quad (2)$$

Nonlinear (NonLin.Reg.). Similar to the impairment-factor structure of ITU-T G.1072 [12], the nonlinear model represents quality as a maximum value discounted by network impairments. Equation (3) models QoE as an exponential decay from the ceiling Q_{Max}, driven by a weighted sum of RTT, PL, and RJ. The exponential form imposes monotonic degradation with increasing impairment and approaches, but does not cross, the MOS floor of 1.

$$x = 1 + (Q_{\text{Max}} - 1)\, e^{-(n_1 \cdot \text{RTT} + n_2 \cdot \text{PL} + n_3 \cdot \text{RJ})} \quad (3)$$

The ceiling is set to the highest observed MOS, Q_{Max}=4.4839, and only the impairment weights n_1, n_2, n_3 are fitted. Fixing the ceiling prevents the nonlinear fit from using an unrealistically high maximum quality to compensate for sparse data at low impairment levels. Because Q_{Max} is nonetheless derived from the data, it is counted as a model parameter (four in total) in the degrees-of-freedom and AICc calculations below.

Range projection. For the two unconstrained models we clip predictions to the admissible range,

$$\text{MOS}_{\text{QoE}} = \max(1, \min(x, 5)) \quad (4)$$

which leaves in-range predictions unchanged and prevents invalid MOS values.

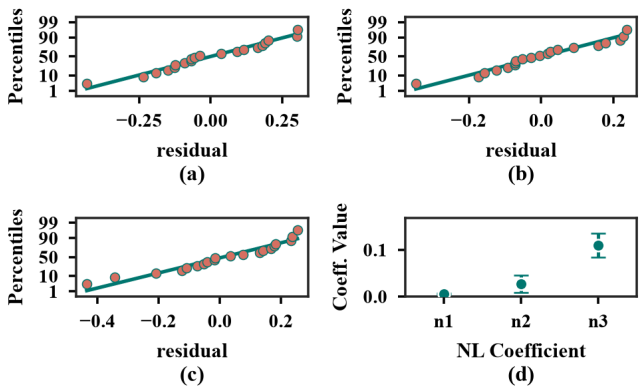

Fig. 1. Residual normal probability plots for (a) Lin.Reg., (b) Poly.Reg., (c) NonLin.Reg., and (d) the 95% confidence intervals of the NonLin.Reg. impairment weights n_1, n_2, n_3.

TABLE III
MODELS' PERFORMANCE ACROSS METRICS.

Model	RMSE	CV (RMSE)	MAE	CV (MAE)	PLCC	R^2	R^2_{Adj}	DF
Lin.Reg.	0.19	0.27	0.16	0.23	0.95	0.91	0.88	15
Poly.Reg.	0.15	0.20	0.12	0.17	0.97	0.94	0.92	14
NonLin.Reg.	0.19	0.23	0.15	0.19	0.95	–	–	16

IV. RESULTS

Models performance was evaluated using five complementary metrics. Prediction accuracy was measured by root mean square error (RMSE) and mean absolute error (MAE), agreement with measured MOS by the Pearson linear correlation coefficient (PLCC), and goodness of fit by the coefficient of determination (R^2) and adjusted R^2. The fitted model coefficients are listed in Table II; the resulting performance metrics are reported in Table III.

A. Fit Validity and Accuracy

Regression analysis assumes approximately normal residuals [13]. The normal probability plots in Fig. 1(a)–(c) show residuals close to the reference line for all three models. Moreover, the Shapiro–Wilk test does not reject residual normality (Lin.Reg. $p = 0.61$, Poly.Reg. $p = 0.50$, NonLin.Reg. $p = 0.17$). For the nonlinear model, Fig. 1(d) reports the 95% confidence intervals of the estimated impairment weights n_1, n_2, and n_3.

Table III shows low prediction error for all three models. The polynomial model gives the lowest error (RMSE = 0.15, MAE = 0.12, $R^2 = 0.94$). The nonlinear model has similar in-sample error to the linear model (RMSE = 0.19 for both), while using fewer parameters and producing range-bounded predictions. R^2 and adjusted R^2 are omitted for the nonlinear model, as these statistics lack a standard interpretation for nonlinear regression. Fig. 2 confirms that this close agreement extends to every individual condition across the RTT, (RTT, PL), and (RTT, RJ) groups. Fig. 3 traces each model's fitted MOS_{QoE} against RTT at fixed impairment levels, exposing differences in functional form that the error metrics alone do not capture.

TABLE IV
AICC-BASED MODEL SELECTION (k: NUMBER OF PARAMETERS).

Model	k	SSR	AICc	ΔAICc
Poly.Reg.	6	0.44	−53.1	0.0
NonLin.Reg.	4	0.70	−52.7	0.3
Lin.Reg.	5	0.70	−48.6	4.5

Generalization is assessed with leave-one-out cross-validation (LOOCV), where each model is repeatedly trained on 19 conditions and evaluated on the omitted condition. LOOCV is appropriate for this dataset because some condition groups, especially packet loss, contain too few samples for representative k-fold splits. The cross-validated errors in Table III exceed the in-sample errors by only 0.04–0.08 MOS.

B. Feature Importance

Beyond predictive accuracy, the practical value of a QoE model to network operators lies in revealing which QoS factor most strongly drives QoE, and hence where provisioning and mitigation effort yields the largest perceptual gain. We quantify this with Shapley values [14], computed as each input's average marginal contribution over the tested range. Fig. 4 shows that RJ has the largest mean absolute Shapley value in all three models (0.40, 0.37, and 0.39 for Lin.Reg., Poly.Reg., and NonLin.Reg., respectively). RTT ranks second (0.27–0.32), and PL has the smallest contribution (0.10–0.14). The same ordering across the three functional forms suggests that the ranking is not specific to one regression specification. This result is consistent with the subjective analysis from Mitra et al. [5], where jitter produced considerable QoE degradation.

C. Model Selection

Because the models differ in parameter count and are not nested, we use the small-sample-corrected Akaike Information Criterion (AICc) [15], which penalizes the number of parameters and, unlike an extra-sum-of-squares F-test, is valid for comparing non-nested models. Table IV reports AICc and the difference from the best model, ΔAICc. The polynomial model has the lowest AICc. The non-linear model is statistically indistinguishable from it ($\Delta\text{AICc} = 0.3 < 2$) while using fewer parameters, whereas the linear model has considerably less support ($\Delta\text{AICc} = 4.5$). Thus the polynomial model is the best-supported fit, with the non-linear model a near-equivalent, range-bounded alternative.

D. Real-Network Assessment

To evaluate the models under measured access-network conditions, we used seven days of measurement data from northern Sweden to five European AWS regions (Frankfurt, Ireland, London, Paris, and Stockholm). One ICMP probe per second was sent over wired, 4G, 5G Non-Standalone, and

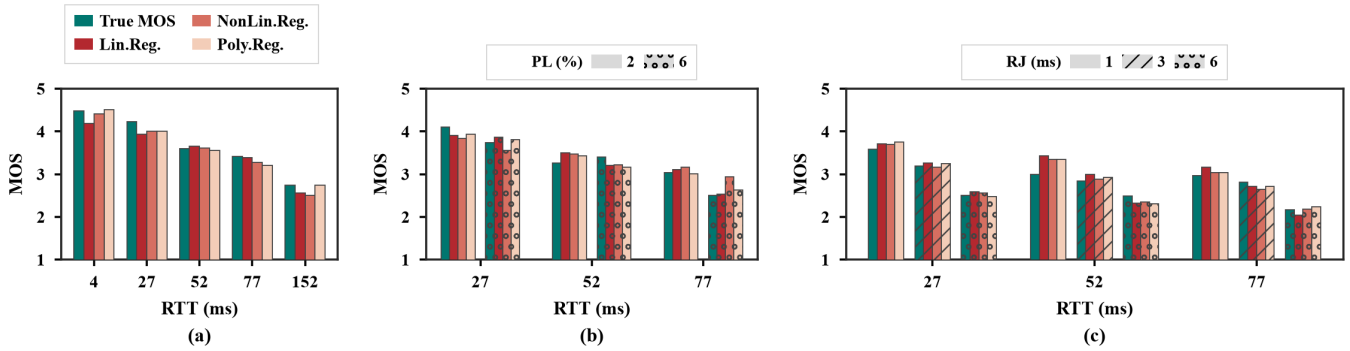

Fig. 2. Predicted versus measured MOS_{QoE} for every condition, grouped by (a) RTT, (b) RTT and PL, and (c) RTT and RJ.

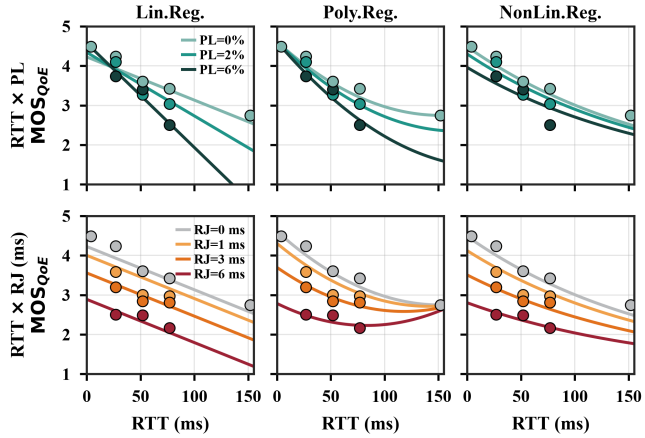

Fig. 3. Fitted MOS_{QoE} response over RTT for each model (columns), sliced by packet loss (top, RJ = 0) and random jitter (bottom, PL = 0); markers are the measured conditions. Curves are drawn unclipped to expose each model's extrapolation behaviour.

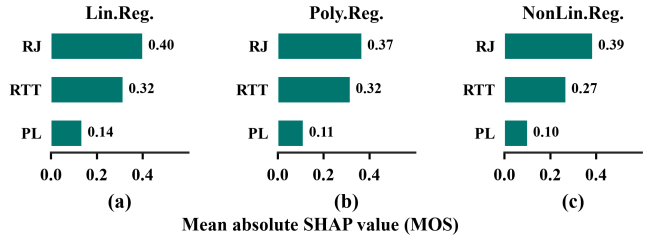

Fig. 4. Shapley feature importance for (a) Lin.Reg., (b) Poly.Reg., and (c) NonLin.Reg.; RJ leads in all three models, then RTT, with PL last.

5G Standalone access networks. The measured RTT, jitter, and packet loss were then used as inputs to the polynomial QoE model. Fig. 5 reports the mean RTT and the jitter as standard deviation of RTT in milliseconds (ms); packet loss was negligible throughout the week.

Across the five datacenters, the predicted MOS_{QoE} ranges from 3.4 to 4.2 for wired, 2.6 to 3.0 for 5G Standalone, and 2.2 to 2.3 for 4G ("Fair"-to-"Good" for wired and "Poor"-to-"Fair" for the mobile networks). 5G Non-Standalone exhibits the highest jitter of the measured networks (RJ ≈ 10 standard deviation from RTT in ms), which exceeds the range observed in the subjective dataset ($\leq$ 6 ms); its QoE therefore lies outside the model's validity domain and is not reported. These results are consistent with the Shapley analysis, in which RJ has the largest contribution to predicted MOS variation.

V. Related Work

Pérez et al. [6] organize QoE assessment methods for immersive communication, and empirical studies examine specific impairments and use cases: Cortés et al. [8] derive delay thresholds for volumetric XR interaction, and Viola et al. [16] evaluate digital humans in immersive communication. For streamed social XR specifically, Cortés et al. [17] propose a QoE model, but predict quality from application-level delay rather than measurable network QoS. Objective QoE prediction from network conditions has instead been pursued in adjacent domains: Rossi et al. [18] fit objective QoE models for cloud gaming over mobile networks, and Tripathi et al. [19] assess the impact of network QoS on metaverse VR experience, while measurement studies characterize social-VR traffic and platforms [20], [21]. The present work builds on the subjective social-XR study of Mitra et al. [5], conducted over controlled RTT, PL, and jitter conditions, whose per-condition MOS provides the basis for the objective models developed here.

VI. Conclusion

This paper developed and validated three objective QoE models for streamed social XR: linear, polynomial, and range-bounded non-linear based on a dataset. All three models satisfied the residual-normality check and yielded low prediction error. The polynomial model attains the lowest error ($R^2 = 0.94$) and the lowest AICc, while the nonlinear model is statistically indistinguishable from it (ΔAICc < 2) with fewer parameters and range-bounded output. A Shapley analysis ranks jitter as the largest contributor to predicted MOS across all three models. Applied to a week of live measurements, the polynomial model yields 3.4–4.2 for wired, 2.6–3.0 for 5G Standalone, and 2.2–2.3 for 4G; 5G Non-Standalone is not

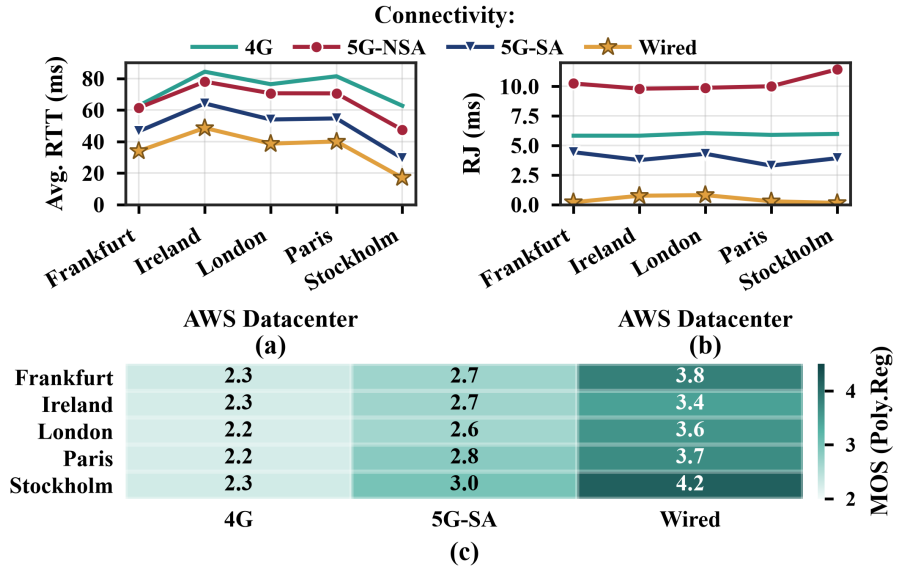

Fig. 5. Measured (a) mean RTT and (b) jitter (ms) to five AWS regions over wired, 4G, 5G Non-Standalone, and 5G Standalone from northern Sweden over seven days, and (c) the resulting polynomial-model MOS_{QoE} for live traffic (shown for 4G, 5G Standalone, and wired).

reported, as its measured jitter lies outside the range observed in the training data.

References

[1] K. Mitra, A. Zaslavsky, and C. Åhlund, "Context-aware QoE modelling, measurement, and prediction in mobile computing systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 920–936, 2015.

[2] H. Wang, H. Dong, and A. El Saddik, "Immersive multimedia communication: State-of-the-art on extended reality streaming," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 21, no. 7, pp. 1–33, 2025.

[3] X. Shen, J. Gao, M. Li, C. Zhou, S. Hu, M. He, and W. Zhuang, "Toward immersive communications in 6G," *Frontiers in Computer Science*, vol. 4, p. 1068478, 2023.

[4] R. Petkova, I. Bozhilov, A. Manolova, K. Tonchev, and V. Poulkov, "On the way to holographic-type communications: Perspectives and enabling technologies," *IEEE Access*, vol. 12, pp. 59 236–59 259, 2024.

[5] K. Mitra, H. S. Rossi, J. Gavrell, and C. Åhlund, "Quality of experience assessment for streamed social extended reality applications over heterogeneous access networks," in *2025 17th International Conference on Quality of Multimedia Experience (QoMEX)*, 2025, pp. 1–7.

[6] P. Pérez, E. Gonzalez-Sosa, J. Gutiérrez, and N. García, "Emerging immersive communication systems: Overview, taxonomy, and good practices for QoE assessment," *Frontiers in Signal Processing*, vol. 2, p. 917684, 2022.

[7] M. Montagud, Á. Egea Benavente, M. Martos, J. Montesa, F. S. Ibáñez, and S. Fernández, "Social eXtended reality (XR) and virtual production: Toward new engaging immersive experiences," in *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX)*. New York, NY, USA: Association for Computing Machinery, 2025, pp. 458–461.

[8] C. Cortés, I. Viola, J. Gutiérrez, J. Jansen, S. Subramanyam, E. Alexiou, P. Pérez, N. García, and P. Cesar, "Delay threshold for social interaction in volumetric extended reality communication," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 7, pp. 1–22, 2024.

[9] M. Nguyen, S. Vats, S. Van Damme, J. van der Hooft, M. T. Vega, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Characterization of the quality of experience and immersion of point cloud videos in augmented reality through a subjective study," *IEEE Access*, vol. 11, pp. 128 898–128 910, 2023.

[10] ITU-T, "Interactive test methods for subjective assessment of extended reality communications," International Telecommunication Union, Geneva, Switzerland, Recommendation ITU-T P.1321, Oct. 2025.

[11] "Principles of subjective test methods for interactive virtual reality (VR) applications," International Telecommunication Union, Geneva, Switzerland, ITU-T Recommendation P.812, May 2024.

[12] ITU-T, "Opinion model predicting gaming quality of experience for cloud gaming services," International Telecommunication Union, Geneva, Switzerland, Recommendation ITU-T G.1072, Jan. 2020.

[13] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 6th ed. Hoboken, NJ, USA: John Wiley & Sons, 2021.

[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.

[15] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York, NY, USA: Springer, 2002.

[16] I. Viola, S. Subramanyam, J. Li, and P. Cesar, "On the impact of VR assessment on the quality of experience of highly realistic digital humans: A volumetric video case study," *Quality and User Experience*, vol. 7, no. 1, p. 3, 2022.

[17] C. Cortés, P. Pérez, and N. García, "Understanding latency and QoE in social XR," *IEEE Consumer Electronics Magazine*, vol. 13, no. 3, pp. 61–72, 2024.

[18] H. S. Rossi, K. Mitra, C. Åhlund, I. Cotanis, N. Ögren, and P. Johansson, "Objective QoE models for cloud-based first person shooter game over mobile networks," in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, 2024, pp. 550–553.

[19] R. D. Tripathi, M. Lyu, and V. Sivaraman, "Assessing the impact of network quality-of-service on metaverse virtual reality user experience," in *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, 2024, pp. 206–213.

[20] A. Alhilal, K. Shatilov, G. Tyson, T. Braud, and P. Hui, "Network traffic in the metaverse: The case of social VR," in *2023 IEEE 43rd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, 2023, pp. 109–114.

[21] R. Cheng, N. Wu, M. Varvello, S. Chen, and B. Han, "Are we ready for metaverse? A measurement study of social virtual reality platforms," in *Proceedings of the 22nd ACM Internet Measurement Conference (IMC)*, 2022, pp. 504–518.