IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 October 2025, accepted 17 November 2025, date of publication 20 November 2025,
date of current version 4 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3635298

RESEARCH ARTICLE

Physiology-Driven User Perception Prediction for Networked Collaborative Virtual Reality

JAVAD SAMERI[1,2], (Graduate Student Member, IEEE), SAM VAN DAMME[1], (Member, IEEE), SUSANNA SCHWARZMANN[3], QING WEI[3], RICCARDO TRIVISONNO[3], (Senior Member, IEEE), FILIP DE TURCK[1], (Fellow, IEEE), AND MARIA TORRES VEGA[2], (Senior Member, IEEE)

[1]IDLab, Department of Information Technology, imec, Ghent University, 9052 Ghent, Belgium
[2]eMedia Research Laboratory, Department of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium
[3]Huawei Technologies, 80992 Munich, Germany

Corresponding author: Javad Sameri (javad.sameri@ugent.be)

This work was supported in part by the Research Foundation Flanders (FWO) WaveVR Project under Grant G034322N, and in part by the FWO Project SENTIENCE under Grant G0A8N25N.

ABSTRACT The growing need for remote immersive collaboration across various societal fields like education, healthcare, and training has given rise to the emergence of Networked Collaborative Virtual Reality (NCVR), where users interact in shared virtual spaces over the network. In such context, the effectiveness of the collaboration will be determined by the quality of the interaction among the users and their perceived quality. However, understanding the user perception in NCVR is complex due to the interplay between network conditions, especially latency and individual user responses. Thus, there is a need for accurate modeling of user perception that could provide fast assessments of the user perception during the collaborative session. Traditionally, assessing user perception has been performed either by subjective studies, where users rank the perceived quality of the content after the session, or by objective metrics, which provide a measure of the quality of the content as compared to the original counterpart. These approaches fall short with NCVR, where not only the quality of the content, but of the interaction and the user's well-being will be key. Physiological signals offer a promising pathway to capture the nuances of the user experience, yet their analysis is challenging due to their inherent complexity and individual variability. This paper makes two key contributions: first, it presents a thorough correlation analysis between user perception and physiological signals. Second, it introduces a novel machine learning method that integrates network performance metrics with physiological data to predict user perception in NCVR environments. The proposed approach was evaluated using an experimental dataset comprising heart activity and network performance data collected from users engaged in a collaborative VR task in the presence of latency-related impairments. Results indicated that combining network data with physiological signals improves prediction accuracy, achieving up to 84% accuracy for latency perception.

INDEX TERMS Collaborative virtual reality, human–computer interaction, network impairments, quality of experience.

I. INTRODUCTION

Recent advancements in Virtual Reality (VR) technology have expanded its applications across many domains, such as gaming, entertainment, healthcare, and education [1]. These applications often provide collaborative experiences, allowing multiple users to interact in shared virtual spaces that require remote collaboration. This increasing need for immersive remote collaboration has driven the emergence of Networked Collaborative VR (NCVR), where

The associate editor coordinating the review of this manuscript and approving it for publication was Antonio Piccinno.

remote users collaborate from different edges of the network.

The effectiveness of NCVR systems depends on users' perceptions of their experiences in shared virtual environments [2]. As they interact over the network, various factors can dynamically influence and shape the user experience. Maintaining positive user perception during an interactive session requires continuous monitoring of user perception across multiple dimensions [3]. Monitoring user perception can be improved by developing accurate models for assessing user Quality-of-Experience (QoE), which is the overall acceptability of users' experience as perceived subjectively. Developing such models enables NCVR systems to optimize parameters, such as network configurations, to improve user engagement and prior work has shown that such models can significantly enhance QoE over the network [4], [5]. However, the variability of network conditions and the subjective nature of user perception present challenges for its modeling [6].

Current NCVR systems predominantly rely on system performance metrics to gain insight into users' perception of VR experience [7], [8], [9]. However, due to VR's immersive nature, which engages a substantial portion of the human sensory system, these estimations often fall short of accurately reflecting the user's experience. This approach tends to overlook critical human factors, as it excludes the nuanced, individual perceptions that are fundamental to understanding the overall user experience [10].

To address this challenge, physiological signals that reflect the user's emotional and perceptual states through measurable responses offer a promising solution [11], [12]. For instance, variations in video quality have been shown to affect brain rhythms [13], and heart rate has been linked to the QoE of media content [14]. Therefore, leveraging these physiological signals can bridge the gap between objective network metrics and the inherently subjective nature of user perception. In this context, latency stands out as one of the critical network impairments shaping interactivity and user perception in VR [15]. Prior research has shown that latency directly influences QoE, making it an appropriate starting point for exploring physiology-driven perception modeling [2]. Accordingly, this study centers on latency-related impairments as a controlled yet representative factor for examining how network conditions influence user physiological responses.

The purpose of this paper is to explore the impact of using physiological data in the assessment of user perception of NCVR. Employing an experimental dataset collected from a collaborative VR task is used, which includes users' subjective assessment as well as physiological indicators in the form of Heart Activity (HA) and network performance metrics under varying latency conditions, this paper contribution is two-folded:

1) A thorough correlation analysis of the user perception and the user's physiology, in particular HA.
2) Based on the correlation results, a machine learning method for modeling user perception by integrating network performance data with physiological markers. By leveraging this dataset, the proposed models' ability to accurately predict user perception in NCVR environments is evaluated.

Our results demonstrate that combining objective and subjective data leads to more reliable models that can dynamically adjust network configurations to enhance the overall user experience.

The remainder of the paper is organized as follows. Section II reviews related work on modeling user perception and the integration of physiological data in VR systems. Section III presents the experimental dataset employed for this research and it explores the link between physiology and user perception in NCVR. Based on these results, we propose our envisioned predictive model to incorporate physiological and network to estimate the user perception of the system (Section IV). In Section V, the previously described methodology is employed for the specific case of the presented dataset, explaining the procedure for modeling user perception. The evaluation results of the envisioned model is presented in Section VI. Section VII discusses the implications of our findings and outlines directions for future research. Finally, Section VIII concludes the paper.

II. RELATED WORK

In this section, an overview of relevant studies regarding modeling and exploring user perception of virtual environments and the integration of physiological signals is provided. Understanding user perception of virtual reality experiences, especially networked collaborative virtual reality (NCVR) environments, is a multifaceted challenge that intersects various disciplines, including network performance, human psychology, design principles, and system engineering. Currently, research in this domain predominantly relies on objective systems such as network performance data metrics to model user perception in VR [8], [9], [16], [17], [18].

For instance, Lee et al. [19] constructed a QoE model for cloud VR gaming, incorporating both objective metrics (e.g., bitrate and network delay) and content-specific factors (e.g., game genre). By employing machine learning modeling, their models demonstrated strong predictive accuracy to model subjective user experiences in virtual environments. In a similar study, Anwar et al. [8] explored the impact of 360-degree video impairments on QoE. They developed a Decision Tree-based model that emphasized factors such as video resolution and interruption frequency as important factors for modeling user experience [20]. In a follow-up work, Anwar et al. [21] conducted a subjective study on 360-degree VR videos, proposing logistic regression and neural network models to predict perceptual quality and cybersickness based on user familiarity and interest, achieving up to 86% prediction accuracy. Similarly, Krogfoss et al. [9] introduced a methodology to correlate key quality indicators specific to XR applications with network key performance indicators, finding that low latency and high throughput are critical for maintaining high QoE in XR.

These approaches, while effective in linking technical performance to user perception, often overlook the inherently subjective and human-centric nature of perception in immersive environments. To gain insights into user perception of their immersive experience, these studies primarily utilized explicit assessments, such as questionnaires and interviews, assuming that users can effectively express their emotions, perceptions, and behaviors. However, these methods are subjective and prone to biases; the language used in questionnaires can be interpreted differently by respondents, and the results often hinge on the participants' self-awareness and current emotional state, which can distort the accuracy of the assessments [21], [22], [23], [24]. Furthermore, advancements in neuroscience have brought attention to the fact that numerous brain functions controlling human emotions and mindsets function subconsciously, outside the scope of consciousness. In contrast to explicit processes, humans cannot verbalize these implicit processes. This realization has driven researchers to explore physiological signals such as Electroencephalograms (EEG), Electrocardiogram (ECG), Blood Volume Pulse (BVP), and Electrodermal Activity (EDA) as tools for capturing these implicit responses and processes. Unlike explicit measures, physiological signals provide direct insights into the user's physical and cognitive states, offering a more objective and continuous assessment of user perception in virtual environments. Studies have demonstrated that these physiological markers are rich in information about the human response to stimuli, making them a promising avenue for enhancing the accuracy of user perception models in NCVR settings [12], [13], [25].

For instance, Hofmann et al. [26] explored the link between subjective emotional arousal and brain activity during an immersive VR experience, utilizing EEG recordings combined with a roller coaster simulation. Their study confirmed that recorded brain activity is associated with higher subjective arousal, effectively decoding arousal states using machine learning techniques. Moreover, Zhang et al. [12] proposed a method leveraging EEG signals to evaluate QoE in VR environments, finding a significant correlation between frontal lobe symetery features and subjective QoE scores. Keighrey et al. [27] compared QoE across Augmented Reality (AR), VR, and tablet-based applications using Heart rate (HR) and Electrodermal activity (EDA) as physiological indicators, revealing increased HR and EDA levels in VR users, which suggested heightened cognitive load and stress during immersive experiences. Mesfin et al. [11] found that physiological metrics, such as eye gaze and heart rate, closely correlated with self-reported QoE. Participants with increased visual attention and elevated heart rates reported higher enjoyment and perceived relevance of the multisensory content, demonstrating that these physiological responses can effectively reflect QoE in immersive media environments.

Similarly, Zheleva et al. [28] investigated the impact of video quality degradation on QoE in VR, highlighting that lower video quality not only reduced subjective ratings of video quality and sensory immersion but also influenced physiological responses, such as lower-alpha activity in EEG recordings, suggesting increased cognitive engagement and reduced immersion under lower quality conditions. Another study by Zheleva et al. [29] explored the impact of multimodal stressors on user experience in VR by measuring EDA in a haunted kitchen scenario. Their findings revealed significant differences in stress responses to audio, visual, and audio-visual stimuli underscoring the potential of integrating physiological measures to create adaptive VR environments that tailor experiences to individual user responses in real-time.

Furthermore, Moharana et al. [30] investigated physiological synchrony in collaborative VR tasks using EDA signals. They found that physiological synchrony was moderately correlated with task performance, such as completion time. Their results suggest that shared background factors, like familiarity and previous technology experience, significantly influence synchrony, highlighting the potential of physiological synchrony as a metric for evaluating collaboration quality in immersive VR settings.

In conclusion, objective metric-based models, though effective in some contexts, often fail to account for the variability in subjective user experiences. These models are typically trained on pre-existing subjective data, which may not align with the current condition or mental state of the user being evaluated. As a result, there is a fundamental disconnection between the model's predictions and the users' current perceptions. This issue underscores the need for models that better integrate real-time user-specific data. On the other hand, consistent correlations between physiological metrics and subjective evaluations of VR experiences in various studies highlight the potential of using physiological measures to assess user perception of VR experiences, highlighting a crucial area for future research to enhance the accuracy and relevance of user perception modeling.

III. EXPERIMENTAL EVALUATION OF CORRELATION OF PHYSIOLOGICAL DATA AND HUMAN PERCEPTION

To explore the correlation between physiological data and human perception, an experimental dataset collected from a collaborative VR task was used. This section begins with a description of the dataset. Next, the processing methods for both objective network data and physiological data are detailed. Finally, the correlation between physiological data, the experimental configurations, and participants' responses to the post-experiment questionnaire is analyzed.

A. DATASET AND EXPERIMENTAL SETUP

To explore the correlation between physiological data and human perception, an experimental dataset collected from a collaborative VR task, as introduced in [2], was utilized.

The dataset includes both network monitoring data between users and the server, as well as the physiological heart activity (HA) of participants during a NCVR experiment. The primary objective of this dataset was to investigate the effect of network latency, which represents one of the most critical impairments influencing user perception in networked collaborative environments [31].

A total of 20 subjects participated in the study, divided into 10 sessions of 2 subjects each. During the visual examination, physiological signals of 3 sessions were excluded due to technical problems, resulting in 7 sessions and 14 participants. The ages of the participants ranged from 22 to 47, with a median of 27.5 and an average of 28.4. Of the participants, 30% identified as female and 70% as male. In accordance with the GDPR, participants were presented with a written informed consent form prior to the session. The form requested permission to collect and process anonymized data, including limited demographic information (e.g., age and gender), for research purposes. It also described potential risks (e.g., cybersickness) and clearly stated participants' rights to withdraw at any time and to request access, rectification, or erasure of their data.

The collaborative experimental task required the participation of two individuals to jointly engage in pizza preparation within a virtual kitchen. The kitchen was purposefully designed with an "H" shaped table layout with tools and ingredients arranged accordingly to foster interactive behavior, positioning two participants at opposing ends, as depicted in Figure 1. For example, preparing the pizza topping required close collaboration between the participants, whereby one would pass the kitchen knife to allow the other to cut the sausage, or alternatively, pass the sausage itself across the table to complete the task. Participants engaged in steps such as combining ingredients, kneading dough, and baking the pizza, each equipped with a Meta Quest 2 VR Head-Mounted Display (HMD) for immersion. They were connected to separate laptops and an individual server managing the virtual environment.

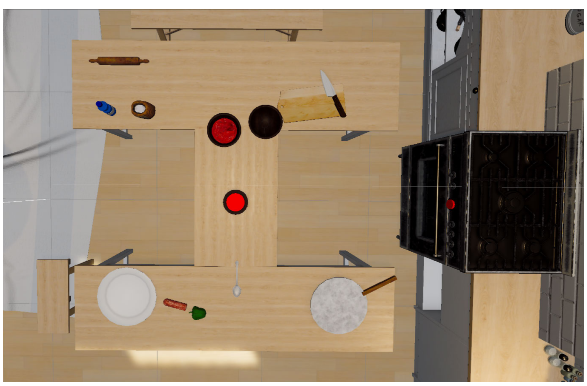


FIGURE 1. Top view of the virtual kitchen environment designed to promote interactive collaboration between participants, with the "H"-shaped table layout requiring them to pass objects.

Traffic analysis and monitoring were conducted using Wireshark, and Photoplethysmogram (PPG) signals were recorded using two Shimmer GSR+[1] devices with a sampling rate of 128 Hz. Four possible scenarios were tested on pairs of participants: Configuration A served as the control with no impairments; Configuration B introduced a 500 ms delay; Configuration C applied a 500 ms burst latency (where communication becomes overwhelmed and is blocked for a specific period) at a 50% occurrence rate; and Configuration D combined both a 500 ms delay and 500 ms burst latency with a 50% occurrence rate. The 500 ms value was chosen in this dataset for its significant impact on user experience in networked VR and alignment with real-world latency, providing a realistic threshold. The four configurations have been depicted in Figure 2. All participants were randomly subjected to each network condition to enable a comprehensive analysis.

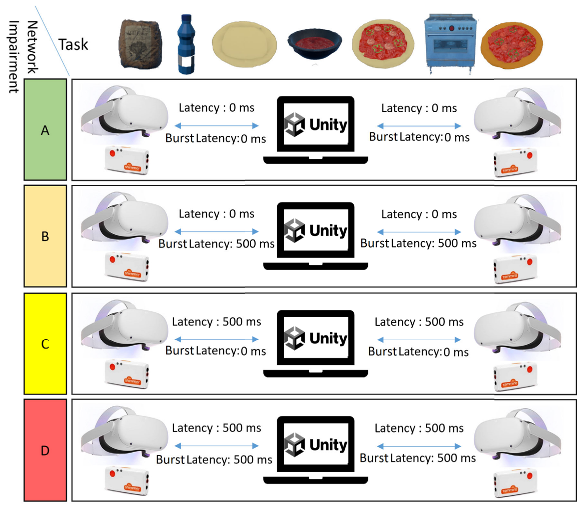


FIGURE 2. Schematic representation of the dataset experimental paradigm with 4 network scenarios denoted as A, B, C, and D.

After completing the experimental tasks, participants underwent subjective evaluation through questionnaires designed to assess their perceived QoE. Participants completed a structured questionnaire immediately after each experimental playthrough. The questionnaire focused on evaluating key aspects of user experience in the VR environment, specifically their perception of network-induced latency, jerkiness, and synchronization. Participants rated their perception of latency across three specific dimensions: general latency perception, latency experienced during collaboration with their partner, and latency perceived when interacting with virtual objects reflecting how network-induced delays between actions and corresponding visual feedback affected their interaction. Throughout this paper, general latency perception is referred to simply as Latency, and Overall Latency is the score calculated by averaging these three dimensions. Additionally, the questionnaire captured participants' perception of jerkiness, referring to sudden or irregular movements within the VR environment. Such jerkiness often arises from burst network

[1] https://shimmersensing.com/product/shimmer3-gsr-unit/

traffic conditions, leading to an uneven and disruptive user experience. Finally, participants assessed synchronization, evaluating the extent of spatial and temporal alignment between their view and their collaborator's perspective in the virtual environment. Each of these criteria was measured using a five-point Likert scale, with a score of 1 indicating the least severity or frequency and a score of 5 indicating the highest. Together, these dimensions provided a comprehensive understanding of subjective quality perceptions crucial to evaluating and predicting user experience in NCVR environments. For a more in-depth information, the reader is referred to these complementary studies [2], [32], which further explains this dataset and NCVR platform and effects of network impairment on user experience.

B. NETWORK (OBJECTIVE) DATA PROCESSING

A set of objective network traffic features was extracted to characterize network conditions that influence user perception. This feature set includes throughput and its statistical descriptors—mean, median, variance, skewness, and kurtosis—offering insights into overall data flow dynamics. Packet size was similarly analyzed using metrics such as average, median, standard deviation, variance, skewness, and kurtosis. Additionally, inter-arrival times of successive packets were computed, providing complementary temporal features including average, median, variance, skewness, and kurtosis. Altogether, these analyses yielded a total of 15 objective features. These features serve as key inputs for subsequent correlation and predictive modeling of user perception. A comprehensive evaluation of these features and their relationship to user experience can be found in [2].

C. HEART ACTIVITY (IMPLICIT SUBJECTIVE) DATA PROCESSING

HA was captured by monitoring BVP fluctuations using a PPG sensor. Due to their convenience, noninvasiveness, and temporal resolution, HA measurements are well-suited to provide implicit subjective indicators of the user's physiological state. However, analyzing HA signals presents challenges, as body movements often introduce noise and artifacts into the BVP signals, making the extraction of heart rhythms more complex [33].

To address these challenges and enhance feature quality, we first opted to record PPG signals from the earlobe instead of the conventional wrist placement. The earlobe is less affected by motion artifacts in ambulatory VR scenarios, where users frequently move their hands to interact with the environment. Second, a dedicated preprocessing pipeline was implemented. This pipeline segments the continuous signal into overlapping 30-second windows with a 50% overlap, analyzing each segment independently.

For each window, the processing begins by filtering the BVP signal through a specified frequency filter between 0.05 Hz and 12 Hz in order to isolate relevant physiological signals and minimize noise that could interfere with HA analysis. Following the filtering process, the signal undergoes a quality check using the signal quality index (SQI) analysis to ensure the integrity and reliability of the data in that specific window [34]. This quality control step is crucial to ensure the Signal-to-Noise ratio (SNR) is within acceptable limits for analysis of the BVP signal.

After confirming signal quality, a comprehensive set of features is extracted from the cleaned signals. These features include various statistical and time-domain measures that capture essential characteristics of the HA, which are critical for subsequent analysis. The extracted features and their descriptions are detailed in Table 1. Finally, biomarkers and features extracted from all acceptable windows are averaged for the span of the recording, synthesizing the data into a form suitable for further analysis or interpretation. Figure 3 provides an overview of the HA signal preprocessing steps.

D. HEART ACTIVITY STATISTICAL ANALYSIS

A series of statistical analyses were conducted to evaluate the impact of different network impairments on participants' HA responses. Overall differences across four configurations were assessed using a Friedman test, with Bonferroni correction, followed by post-hoc Wilcoxon signed-rank tests for pairwise comparisons. Significant differences ($p < 0.05$) were observed in several physiological features, including Tsp, Tdp, and Tswx (x = 10, 25, 33, 50, 66, 75, 90). Figure 4 depicts the first four significant features. Notably, most of these features were related to the systolic wave in the BVP signal [35]. The observed significant changes in Tsp, Tdp, and Tswx suggest that different network latencies could lead to alternation of the cardiovascular dynamics of users, likely due to increased cognitive load, stress, or physical activity [36]. A noticeable decrease in Tsp and Tswx measurement can be observed as the network latency deteriorates from Configuration A without network impairment to Configuration D with 500ms delay and 500ms burst delay with a 50% chance at the same time. These changes in the systolic wave could be a reflection of how network deterioration in NCVR could have an impact on user experience leading to changes in nervous system activity, altering the PPG signal morphology [37].

Additionally, participants' self-reported levels of Perception of Latency, Overall Latency, Jerkiness, and De-Synchronization, were categorized into binary groups for further analysis using the Mann-Whitney U test. For the perception of Latency, AUCdia, Tsys, and Tdp emerged as significant. For the perception of Overall Latency, which compromised the average reported level of Latency, Latency perceived in collaboration, and Latency perceived in object interaction, Tsys emerged as the only significant feature. For the perception of Jerkiness, AuCsys, Tsys, and AUCpi emerged as significant. For the De-Synchronization level, which is the level of de-synchronization subject perceived between themselves and their collaborator, features such as Tdwx (x = 10, 25, 50), Tpwx (x = 10,25,33), Interbeat interval (IBI), Tpi, and HR have emerged significantly.

TABLE 1. Biomarkers derived from the BVP signal and their descriptions.

Biomarker	Description
HR Heart Rate	A straightforward measurement of heartbeats per minute.
IBI Inter-Beat Interval	The time interval between successive heartbeats for assessing Heart Rate Variability.
Tpi Pulse Interval	Time between the beginning and the end of a pulse wave.
Tpp Peak-to-Peak Interval	Time between two consecutive systolic peaks of pulse waves.
Tsys Systolic Time	Time from the pulse onset to the dicrotic notch (point after the systolic peak).
Tdia Diastolic Time	Time from the dicrotic notch to the end of the pulse wave.
Tsp Systolic Peak Time	Time from the pulse onset to the systolic peak (Also known as crest time (CT)).
Tdp Diastolic Peak Time	Time from the pulse onset to the diastolic peak.
ΔT Time Delay	Time delay between the systolic peak and diastolic peak of the pulse wave.
Tswx Systolic Width at x%	Width of the pulse at x% of the systolic peak amplitude, measured between pulse onset and systolic peak.
Tdwx Diastolic Width at x%	Width of the pulse at x% of the systolic peak amplitude, measured between systolic peak and pulse offset.
Tpwx Pulse Width at x%	Total width of the pulse, combining the systolic and diastolic widths at x% of the systolic peak amplitude.
Asp Systolic Peak Amplitude	Amplitude difference between the pulse onset and the systolic peak.
Adn Dicrotic Notch Amplitude	Amplitude difference between the pulse onset and the dicrotic notch.
Adp Diastolic Peak Amplitude	Amplitude difference between the pulse onset and the diastolic peak.
Aoff Pulse Onset Amplitude	Amplitude difference between the pulse onset and the pulse offset.
AUCpi Area Under Curve-Pulse Interval	Area under the pulse wave between the onset and the offset (pulse interval area).
AUCsys Area Under Curve-Systolic	Area under the pulse wave from the onset to the dicrotic notch (systolic area).
AUCdia Area Under Curve-Diastolic	Area under the pulse wave from the dicrotic notch to the offset (diastolic area).
BR Breathing Rate	Estimated breathing rate extracted from BVP signal.

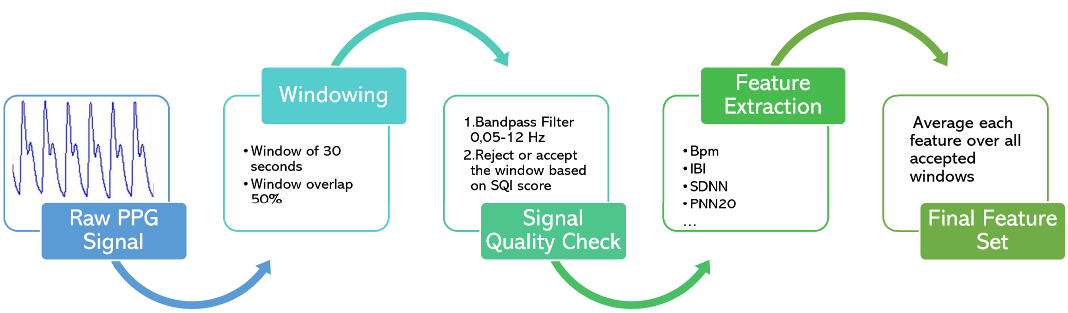
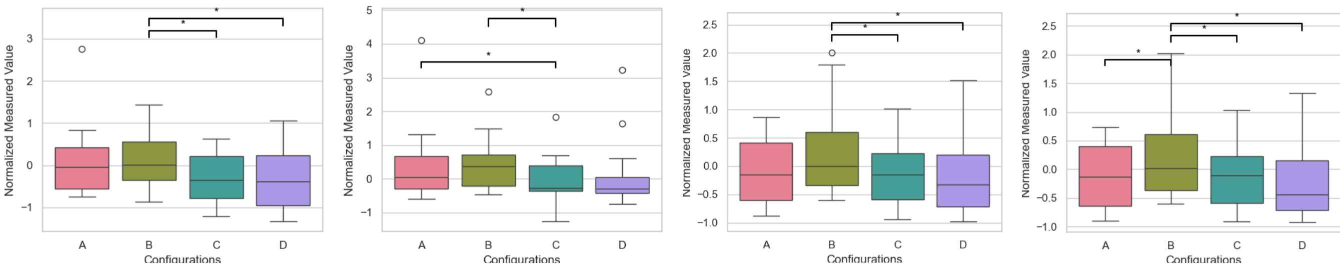


FIGURE 3. Overview of the HA processing steps.

Significant features are depicted in Figure 5. Comparing the significant features and their distributions for Latency and Jerkiness criteria reveals distinct physiological patterns among participants who perceived these impairments in the NCVR environment. For participants who reported experiencing Latency and Jerkiness, there is a notable reduction in the average Tsys and a smaller systolic area, specifically AUCdia in Latency and AUCsys in Jerkiness. These patterns are consistent with responses commonly associated with stress or increased physical effort suggesting



(a) Systolic Peak Time (Tsp) (b) Diastolic Peak Time (Tdp) (c) Systolic Width at 25% (Tsw25) (d) Systolic Width at 50% (Tsw50)

FIGURE 4. Group comparison of the first four significant features: (a) Tsp, (b) Tdp, (c) Tsw25, and (d) Tsw50. These features are primarily related to the systolic wave in the BVP signal. Significant differences among configurations are denoted as * ($p < 0.05$) based on post-hoc Wilcoxon signed-rank tests.

that the perception of Latency and Jerkiness may lead to inducing stress-like symptoms in users or elicit more intense physical activity. Nevertheless, given the present design and dataset, these relationships are best interpreted as associative rather than causal.

In contrast, an intriguing pattern emerges when examining the physiological measurements related to the perception of De-Synchronization. Unlike Latency and Jerkiness, physiological stress markers tend to decrease among participants who perceived the environment as unsynchronized between collaborators. Specifically, in De-Synchronized conditions, the average HR is lower compared to those who reported that they did not perceive De-Synchronization, and the markers Tdw10 and Tpw10, which are associated with the emergence of systolic and diastolic peaks, are significantly higher in the group perceiving De-Synchronization. This contrasting response suggests that users who perceived De-Synchronization may exhibit a lower physiological stress response than those who experienced latency and jerkiness. One possible explanation is that the nature of De-Synchronization as an impairment may lead to a different type of cognitive, emotional, or even Physical reaction, possibly characterized by reduced stress.

IV. ENVISIONED HUMAN CENTRIC NETWORK CVR PLATFORM

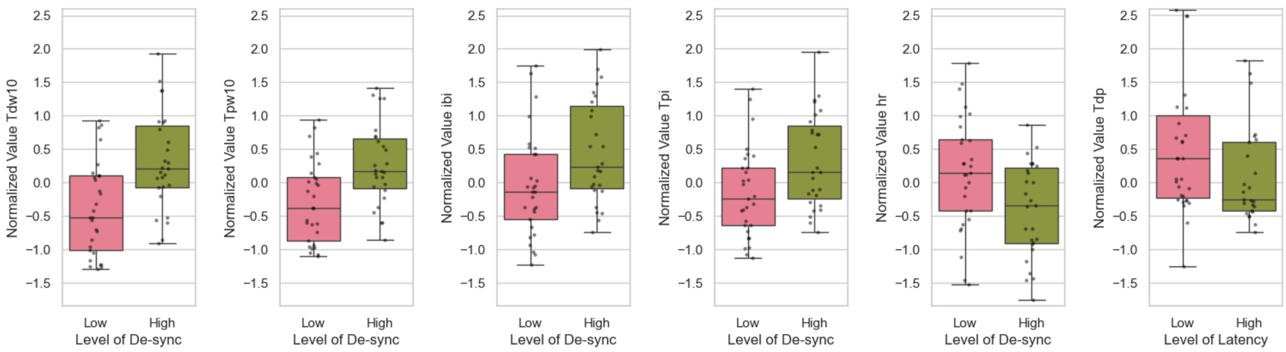
The envisioned NCVR configuration aims to integrate user perceptual modeling designed to provide feedback to the system management, thereby enhancing the overall user experience within NCVR environments. This approach recognizes that user perception is influenced not only by the technical performance of the system but also by individual subjective experience. These subjective experiences are reflected in physiological responses, underscoring the importance of incorporating human-centric factors into the assessment of user perception.

Figure 6 shows the architecture of the envisioned human-centric NCVR system. This architecture comprises three main components: the *server infrastructure*, *the core network*, and the *client infrastructure*.

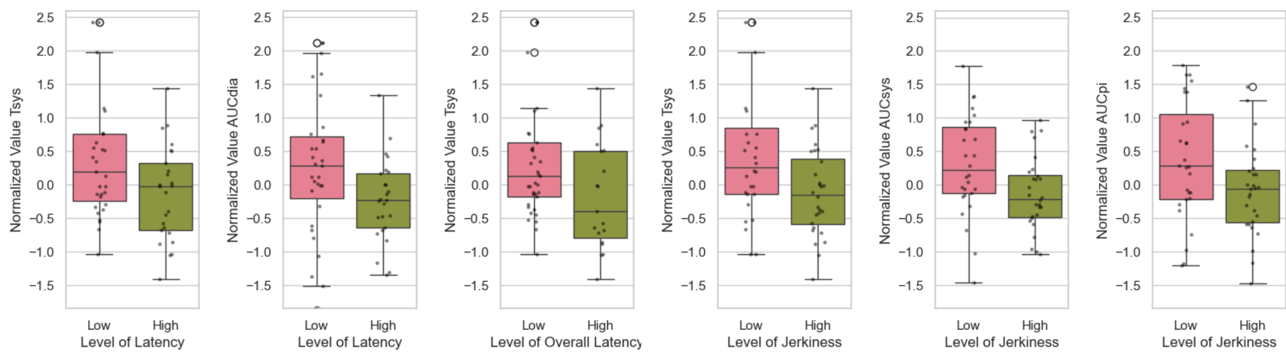
The *server* manages the shared virtual environment and interacts with clients over the network, maintaining the synchronization of the shared space. Simultaneously, data captured from *clients*, which includes network performance metrics and physiological state information, is transferred to the *server*, enabling the system to monitor and assess user perception in real-time or near real-time.

Within the client infrastructure, *network monitoring* and *physiological state monitoring* modules continuously collect data. The *network monitoring* component captures network key performance metrics, which then are processed to form objective data inputs. The *physiological state monitoring* component collects data reflecting the user's physiological responses, such as heart rate variability, galvanic skin response, and other biometric indicators, which are indicative of the user's emotional and perceptual state. These signals undergo *data processing* and *feature extraction* to provide a comprehensive set of inputs representing the user's implicit subjective response. The processed data from both the server and client infrastructures are then fed into the *user perception model*. This model integrates the objective network metrics with the physiological markers, creating a predictive model of user perception that maps the combined inputs to a scale reflective of perceived experience. By continuously estimating the user perception of the system, the model can provide feedback to adjust network configurations dynamically, such as bandwidth allocation or latency management, thus enhancing the overall user experience.

This approach is based on the understanding that user experience in NCVR environments is shaped by a complex interplay between technical performance, individual perception, and interaction within the virtual space. Given the multifaceted nature of user perception, a purely objective model would be insufficient, as it fails to account for individual variability in responses to identical conditions. The integration of physiological signals allows the model to capture these personal nuances, providing a more holistic and accurate prediction of user perception. More specifically, objective-based models that rely solely on objective metrics can predict user perception without the need for a human subject, treating user experience as a static factor once the model is trained. These models could continue to estimate perception as if it exists independently of the human subject, ignoring the inherently subjective nature of user experience.



(a) Diastolic Width at 10% (Tdw10), De-Synchronization (b) Pulse Width at 10% (Tpw10), De-Synchronization (c) Inter-Beat Interval (IBI), De-Synchronization (d) Pulse Interval (Tpi), De-Synchronization (e) Heart Rate (HR), De-Synchronization (f) Diastolic Peak Time (Tdp), Latency



(g) Systolic Time (Tsys), Latency (h) Area Under Curve Diastolic (AUCdia), Latency (i) Systolic Time (Tsys), Overall Latency (j) Systolic Time (Tsys), Jerkiness (k) Area Under Curve Systolic (AUCsys), Jerkiness (l) Area Under Curve Pulse Interval (AUCpi), Jerkiness

FIGURE 5. Identified physiological features significantly associated with user-reported latency, jerkiness, and De-Synchronization in NCVR environments.

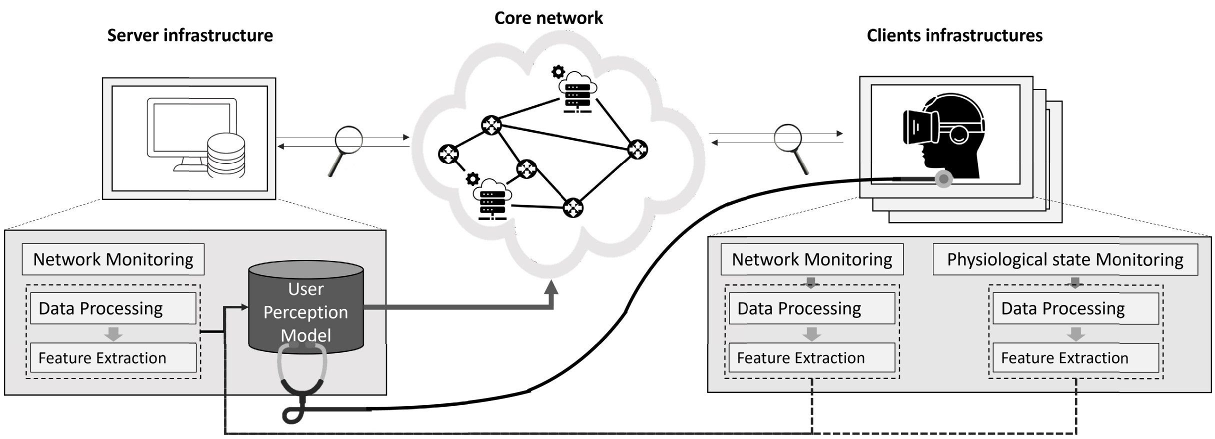


FIGURE 6. Predictive perception modeling architecture.

In contrast, this proposed modeling technique requires implicit subjective data from the user to produce meaningful outputs that reflect the user's current state. This approach ensures the model remains connected to the user's dynamic experience, bringing us closer to accurately modeling the subjective experience of users in NCVR systems and opens up possibilities for personalization of the NCVR system.

V. USER PERCEPTION MODELING IN THE NETWORK COLLABORATIVE VIRTUAL REALITY PIZZA BAKING ENVIRONMENT

User perception in this NCVR environment was modeled based on four key criteria derived from subjective questionnaires: Latency, Overall Latency, Jerkiness, and De-Synchronization. These criteria were selected due to their significant impact on user experience in VR contexts. The models were designed to accurately predict users' perception of the immersive environment, with outputs that closely aligned with the subjective evaluations reported by the users.

The modeling approach utilized two types of input data: objective data from network communication metrics and implicit subjective measurements, namely the HA data. The structure of the model's inputs and outputs is illustrated in Figure 8.

The full end-to-end pipeline for predictive modeling, including all processing steps, is summarized in Figure 7. The process begins with *data collection*, *preprocessing* and *normalization*, during which data is formatted and cleaned to remove noise and artifacts. This step is followed by *feature extraction*, where relevant characteristics are derived from the data. Details of *feature extraction* from network traffic and HA were explained in Section III. A critical challenge in modeling subjective data, such as questionnaire responses, is the class imbalance, which can skew model performance towards the majority class. To mitigate this, the *Synthetic Minority Oversampling Technique (SMOTE)* was employed to generate synthetic samples for the minority class, ensuring a balanced class distribution [38]. This oversampling was applied only to the training set to preserve the real-world distribution of the validation and test sets.

High-dimensional data with correlated features, particularly those extracted from PPG signals, posed a risk of overfitting due to their shared morphological characteristics (e.g., features in the Tdwx and Tpwx series). To address this, *Principal Component Analysis (PCA)* was used for dimensionality reduction, transforming the data into orthogonal components that capture the most significant variance while removing redundancy. This approach enhances the model's generalizability.

Following feature extraction and dimensionality reduction, *hyperparameter optimization* was conducted to refine the predictive models. Three machine learning algorithms—Decision Tree (DT), Support Vector Classifier (SVC), and XGBoost (XGB)—were evaluated for each criterion to produce binary predictions of user perception, distinguishing between the presence and absence of perceived issues as reported by users such as Latency, Overall Latency, Jerkiness, Sync. The reported five-point Likert scores were converted into binary labels to reduce label noise in a small sample and to mitigate class imbalance across participant responses. This binarization was applied solely within the proof-of-concept scope of the current dataset to facilitate a stable demonstration of the modeling framework. However, this conversion does not constrain the general applicability of the proposed framework, which remains label-agnostic and can be extended to ordinal or continuous targets with larger datasets. Adapting the model to generate predictions on a more fine-grained scale would follow the same methodological pipeline, requiring only the availability of appropriately scaled ground-truth data.

Furthermore, a grid search strategy was employed for *hyperparameter* tuning, systematically exploring a range of values to identify the optimal configuration for each model. The Maximum Depth parameter for DTs was varied (values: 2, 4, 5, 8, 10). For SVC, different kernel types (RBF, linear) and cost parameters (0.1, 1, 10) were explored, while the number of estimators (10, 50, 100) and Maximum Depth were adjusted for XGB.

Hyperparameter optimization was executed within a *5-fold cross-validation* loop. Moreover, the performance of the trained models was evaluated using *leave-one-out cross-validation*, where each data point was iteratively used as a test sample while the remainder served as the training set. These cross-validation loops provide a comprehensive assessment of the modeling approach's predictive accuracy and overall performance.

VI. RESULTS AND ANALYSIS

A. USER PERCEPTION MODELING PERFORMANCE

The performance of the models, evaluated in terms of Balanced Accuracy (BAC) as shown in Equation 1, is summarized in Table 2. BAC is a metric that accounts for class imbalance by averaging the accuracy obtained in each class, providing a more comprehensive measure of performance than overall accuracy alone. It is computed as:

$$\mathrm{BAC} = \frac{1}{2}\left(\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} + \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}\right) \tag{1}$$

Here, TP, FN, TN, and FP denote true positives, false negatives, true negatives, and false positives, respectively. To assess the efficacy of employing combined data on user perception modeling, two additional sets of models were trained, each using a single modality: one set trained exclusively on objective network features and another set trained solely on physiological features.

For the Latency perception criterion, models utilizing combined features consistently outperformed those using either objective or physiological features alone. Specifically, the SVC model achieved the highest BAC of 0.78, illustrating that the integration of network metrics with physiological responses substantially improves the model's ability to predict perceptions related to latency. Similarly, for the Overall Latency criterion, models with combined features showed superior performance, with the SVM model achieving the highest BAC of 0.84. This indicates that perceptions of Overall Latency are more accurately represented when both objective and physiological data are integrated. Moreover, in the De-Synchronization criterion, the combined approach continued to excel, with both SVC and DT models reaching the highest BAC scores of 0.79. These results suggest that perceptions of latency and synchronization are significantly influenced by both objective and subjective factors, which are most effectively captured through a combined modeling approach, yielding high accuracy in estimates.

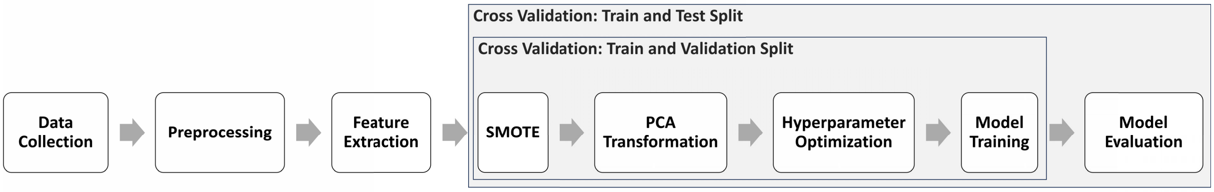


FIGURE 7. End-to-end machine learning pipeline for user perception prediction.

TABLE 2. Results for the perception modeling approach based on different features and each perception criterion.

perception Criterion	Classifier	Objective Network Features BAC	Physiological Features BAC	Combined Features BAC
Latency	DT	0.63	0.58	**0.78**
	SVM	0.57	0.55	**0.79**
	XGB	0.57	0.54	**0.68**
Overall Latency	DT	0.54	0.49	**0.59**
	SVM	0.56	0.57	**0.84**
	XGB	**0.56**	0.52	0.52
Jerkiness	DT	**0.65**	0.47	0.43
	SVM	0.36	**0.62**	0.34
	XGB	**0.62**	0.56	0.40
Sync	DT	0.44	0.51	**0.77**
	SVM	0.55	0.45	**0.79**
	XGB	0.43	0.56	**0.60**

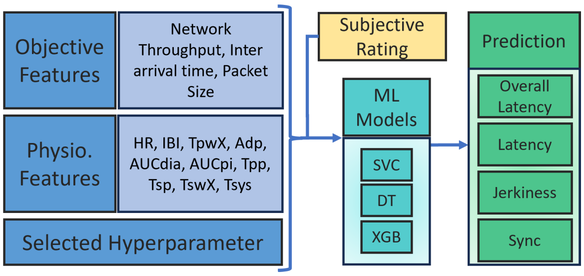


FIGURE 8. Input-output structure of the user perception prediction model.

For the Jerkiness criterion, the highest BAC score of 0.65 was achieved by the DT model using objective network features. SVM and XGB classifiers exhibited similar trends, with objective features generally outperforming other modalities. Models combining objective and physiological data did not surpass the performance of objective-only models, despite having access to a broader range of inputs. This counterintuitive outcome may be due to the increased dimensionality and complexity introduced by physiological features, limiting the models' ability to generalize effectively given the small dataset size.

Overall, the analysis suggests the efficiency of using combined models that integrate both objective network features and physiological measurements in predicting user perception. SVM and DT appeared to be the most effective classifiers across different criteria, demonstrating the potential of machine learning models to provide nuanced insights into user perception in NCVR.

B. NEAR REAL-TIME USER PERCEPTION ESTIMATION

As real-time adaptability is critical for enhancing user experience in networked collaborative VR systems, it is essential to ensure that user perception models can operate within resonable temporal constraints. To this end, all computational and timing analyses were performed on a Dell laptop equipped with a 12th Gen Intel Core i7-1265U CPU and 16 GB of RAM. In this part, we analyze the computational and time complexity of the developed perception models to gain insights into their computational requirements. Understanding the computational and time complexity of different classifiers and feature sets (objective, physiological, and combined) is essential for assessing their suitability for near real-time or real-time applications. The time complexity of the models evaluated in section VI-A is presented in Table 3.

The results indicate that while the BAC performance generally improves when combining objective and physiological features, this enhancement often leads to increased execution times. Notably, models that utilize physiological data alone exhibit significantly longer execution times compared to those using objective features, with combined features demonstrating the highest complexity and variance.

For example, in the Latency and Overall Latency criteria, SVM demonstrates high BAC (0.79, 0.84) scores with combined features; however, this improvement is accompanied by increased execution times, which may pose challenges for near real-time deployment. In contrast, DT models, although showing relatively lower BAC (0.78, 0.59) performance, particularly with physiological and combined features, offer

TABLE 3. Computational and time complexity for different features. The measurements are in seconds and represent the mean time taken for the entire pipeline to run. The values in parentheses are the variances of these measurements.

Criteria	Classifiers	Objective Features	Physiological Features	combined Features
Latency	DT	0.1263 (0.000251)	4.410660 (0.099001)	7.722871 (2.950464)
	SVM	0.108920 (0.000092)	4.306285 (0.537582)	5.403857 (0.578399)
	XGB	0.190310 (0.000082)	6.049196 (1.129962)	7.256219 (4.75488)
Overall Latency	DT	0.24241(0.000891)	6.241185(1.183482)	7.683396 (3.909704)
	SVM	0.20916(0.001148)	6.320907 (0.742440)	7.426721 (1.533936)
	XGB	0.31172(0.001494)	4.857800(0.037570)	6.903918 (0.211455)
Jerkiness	DT	0.106360 (0.000024)	4.491127 (0.478598)	6.576661 (1.074899)
	SVM	0.114672 (0.000186)	4.146181 (0.114579)	6.357893 (1.814998)
	XGB	0.170490 (0.000032)	5.187194 (0.760949)	6.622808 (1.642741)
De-syncronization	DT	0.118680 (0.000122)	5.520509 (2.06434)	5.327630 (0.510015)
	SVM	0.103210 (0.000063)	4.611868 (0.523546)	7.194816 (4.038144)
	XGB	0.177420 (0.000165)	4.538149 (0.271437)	5.474385 (0.516856)

faster execution times, making them a suitable option in scenarios where computational efficiency is prioritized over maximum accuracy.

In the case of jerkiness detection, DT achieves the fastest execution time with objective features (0.106360 s) and also the highest BAC (0.65). For user De-Synchronization detection, DT (BAC = 0.77) provides comparable performance to SVM (BAC = 0.79) but with lower execution time when using combined features.

Overall, the combined analysis of BAC performance and execution time underscores the importance of selecting appropriate classifiers and feature sets based on specific application requirements. SVM tends to offer superior performance compared to DT but incurs the slowest execution times with the highest variances.

VII. DISCUSSION

The critical need for innovative methods to estimate the user perception in NCVR is addressed in this paper. Significant insights regarding the integration of objective and explicit subjective modalities and the implications of real-time processing demands are revealed through the exploration of perception modeling. Together, these observations motivate more adaptive and personalized perception management systems capable of dynamically responding to the unique needs and experiences of individual users. Initially, through the literature and experimental study of this paper, it was observed that objective modalities alone provide a decent estimation of some perception criteria. Notably, they facilitate near real-time estimation (approximately 100 ms), suggesting their viability for real-time perception assessment. Furthermore, models based solely on physiological metrics are found to underperform in comparison and require increased computation time. However, the inclusion of human-centric components via physiological inputs with objective metrics, despite increasing processing times, is shown to enhance perception estimation. This improvement necessitates a critical evaluation of whether the enhanced performance justifies the additional computational overhead. The value of this trade-off depends on the specific real-time requirements and system expectations of the NCVR platform since each application may define 'real-time' differently based on its operational needs and the user's interaction with the system.

Previous research has shown that physiological signals, such as heart activity, capture the emotional responses of users [39]. However, these signals are not well-suited to distinguish the direct effect of transient network impairments, such as latency spikes, lags, or glitches. We hypothesize that the distinctive information captured by our models is the manifestation of these impairments only after they have been processed perceptually or cognitively through the nervous system. Thus, providing information into our model is temporally lagged and has a slower rate compared to objective metrics. Furthermore, we acknowledge that the current physiological dataset is limited to earlobe PPG thus future work should explore multiple modalities such as electroencephalogram (EEG), EDA, and Functional near-infrared spectroscopy (fNIRS) to provide a more robust and comprehensive understanding of user perception.

Based on this insight, we can envision a hierarchical modeling approach that accounts for the differing frequencies and nature of data inputs. In this framework, one level would manage slower-changing user states derived from physiological signals, while another would address high-frequency, transient changes from objective metrics.

This separation allows the model to handle data inputs at their respective frequencies, retaining critical temporal information that might be lost when combining them directly, a limitation often encountered with traditional modeling techniques such as the ones employed in this paper.

Regarding the near-real-time assessment of the user perception model, additional latency factors beyond data collection must be considered. Specifically, the latency associated with transporting processed physiological information from the user at the network edge to the server introduces further delay, particularly under network congestion conditions.

Additionally, the near real-time processing pipeline, initially tested on a single thread, has shown potential for performance enhancement through multi-threading, which could significantly reduce processing times. Follow-up experiments will measure end-to-end latency under multi-threaded feature extraction and inference, profiling CPU/GPU utilization and identifying feature subsets that preserve accuracy. However, it is important to consider the computational cost when selecting features, as not all features can be calculated with the same efficiency.

Moreover, in terms of model performance, SVM and DT demonstrated superior accuracy compared to XGB. The relatively lower performance of XGB could be attributed to the limited dataset size, which may cause overfitting in complex models such as XGB. This suggests that larger datasets could better harness XGB's potential, enabling its strengths in multi-threading and ensemble capabilities to complement DT models, thus leveraging DT models' faster execution times. Accordingly, with an extensive dataset that includes a larger number of participants and a more diverse range of network impairments, we expect that more sophisticated models to benefit from richer training signals and better regularization, potentially outperforming simpler baselines while preserving near–real-time inference.In future work, we plan to incorporate additional pairs of participants and sessions to explore advanced machine learning algorithms and improve model generalization across users.

Moreover, the models performed best for the Latency and Overall Latency criteria, which could be attributed to the dataset's emphasis on latency and burst latency impairments. This suggests that the model's effectiveness could be linked to the types of impairments emphasized during experimentation. The exploration of broader scenarios and diversifying the impairments in dataset will be crucial in refining model accuracy and reliability, especially in real-world applications where network conditions can vary unpredictably. By integrating a wider spectrum of network impairments, future research can build upon the current framework to develop a more generalized and adaptable user perception assessment approach, ensuring that both technical performance and user perception are comprehensively addressed in NCVR environments.

VIII. CONCLUSION

In conclusion, this paper underscores the potential of integrating objective and implicit subjective data streams for user perception modeling in NCVR environments. Statistical analysis suggested that physiological features were significantly affected by effect of network impairments in user perception. Furthermore, the proposed combined models, building on these statistical insights, achieved significant performance, particularly in predicting Latency and Overall Latency perception with accuracies of 0.79 and 0.84, respectively, demonstrating their effectiveness in the assessment of user experience. As NCVR technology evolves, continuous refinement of these models will be essential to meet the increasing demands of near real-time system performance and user satisfaction. Future research should focus on broadening the scope of network impairments studied, incorporating additional physiological signals, and further optimizing model efficiency to align with real-world NCVR applications.

REFERENCES

[1] A. Wexelblat, *Virtual Reality: Applications and Explorations*. New York, NY, USA: Academic, 2014.
[2] S. Van Damme, J. Sameri, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, and M. Torres Vega, "Impact of latency on QoE, performance, and collaboration in interactive multi-user virtual reality," *Appl. Sci.*, vol. 14, no. 6, p. 2290, Mar. 2024.
[3] H. Dong and J. S. A. Lee, "The metaverse from a multimedia communications perspective," *IEEE MultimediaMag.*, vol. 29, no. 4, pp. 123–127, Oct. 2022.
[4] J. Santos, S. Van Damme, J. Sameri, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, and M. T. Vega, "Leveraging user perception for 6G edge-cloud orchestration of networked extended reality," *IEEE Commun. Mag.*, early access, Jun. 2, 2025, doi: 10.1109/MCOM.003.2400594.
[5] J. Sameri, J. Santos, S. Van Damme, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, and M. Torres Vega, "Reinforcement learning-based orchestration of XR applications in distributed 6G cloud infrastructures," in *Proc. 21st Int. Conf. Netw. Service Manage. (CNSM)*, Oct. 2025, p. 7.
[6] J. Ruan and D. Xie, "Networked VR: State of the art, solutions, and challenges," *Electronics*, vol. 10, no. 2, p. 166, Jan. 2021.
[7] S. Vellingiri and P. Balakrishnan, "Modeling user quality of experience (QoE) through position discrepancy in multi-sensorial, immersive, collaborative environments," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 296–307.
[8] M. S. Anwar, J. Wang, S. Ahmad, W. Khan, A. Ullah, M. Shah, and Z. Fei, "Impact of the impairment in 360-degree videos on users VR involvement and machine learning-based QoE predictions," *IEEE Access*, vol. 8, pp. 204585–204596, 2020.
[9] B. Krogfoss, J. Duran, P. Perez, and J. Bouwen, "Quantifying the value of 5G and edge cloud on QoE for AR/VR," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–4.
[10] A. Haj-Bolouri, "The experience of immersive virtual reality: A phenomenology inspired inquiry," *Commun. Assoc. Inf. Syst.*, vol. 52, pp. 782–814, 2023.
[11] G. Mesfin, N. Hussain, E. Kani-Zabihi, A. Covaci, E. B. Saleme, and G. Ghinea, "QoE of cross-modally mapped mulsemedia: An assessment using eye gaze and heart rate," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 7987–8009, Mar. 2020.
[12] Y. Zhang, Y. Su, and X. Sun, "A QoE physiological measure of VR with vibrotactile feedback based on frontal lobe power asymmetry," *IEEE Trans. Multimedia*, vol. 26, pp. 2932–2942, 2024.
[13] E. Kroupi, P. Hanhart, J.-S. Lee, M. Rerabek, and T. Ebrahimi, "User-independent classification of 2D versus 3D multimedia experiences through EEG and physiological signals," in *Proc. 8th Int. Workshop Video Process. Quality Metrics Consum. Electron.*, 2014, p. 6.
[14] S. Vijayakumar, R. Flynn, P. Corcoran, and N. Murray, "BiLSTM-based quality of experience prediction using physiological signals," in *Proc. 14th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Sep. 2022, pp. 1–4.
[15] C. Cortés, P. Pérez, and N. García, "Understanding latency and QoE in social XR," *IEEE Consum. Electron. Mag.*, vol. 13, no. 3, pp. 61–72, May 2024.
[16] M. Shahid Anwar, J. Wang, S. Ahmad, A. Ullah, W. Khan, and Z. Fei, "Evaluating the factors affecting QoE of 360-degree videos and cybersickness levels predictions in virtual reality," *Electronics*, vol. 9, no. 9, p. 1530, Sep. 2020.
[17] P. Pérez, "Exploring the realverse: Building, deploying, and managing QoE in XR communications," in *Proc. ITU Kaleidoscope- Extended Reality-How Boost Quality Exper. Interoperability*, Dec. 2022, pp. 1–11.
[18] G. Kougioumtzidis, A. Vlahov, V. K. Poulkov, P. I. Lazaridis, and Z. D. Zaharis, "Deep learning-aided QoE prediction for virtual reality applications over open radio access networks," *IEEE Access*, vol. 11, pp. 143514–143529, 2023.

[19] K.-Y. Lee, J.-W. Fang, Y.-C. Sun, and C.-H. Hsu, "Modeling gamer quality-of-experience using a real cloud VR gaming testbed," in *Proc. 15th Int. Workshop Immersive Mixed Virtual Environ. Syst.*, Jun. 2023, pp. 12–17.

[20] K. Brunnström, S. A. Beker, K. De, A. Dooms, S. Egger, M.-N. Garcia, T. Hossfeld, S. Jumisko-Pyykkö, C. Keimel, and M.-C. Larabi, "Qualinet white paper on definitions of quality of experience qualinet white paper on definitions of quality of experience output from the fifth qualinet meeting, novi sad," in *Serbia, European Network on Quality of Experience in in Multimedia Systems and Services (COST Action IC 1003)*, no. 26, Mar. 2013. [Online]. Available: https://researchportal.vub.be/en/publications/qualinet-white-paper-on-definitions-of-quality-of-experience-qual/

[21] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad, and Z. Fei, "Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148084–148099, 2020.

[22] J. Marín-Morales, C. Llinares, J. Guixeres, and M. Alcañiz, "Emotion recognition in immersive virtual reality: From statistics to affective computing," *Sensors*, vol. 20, no. 18, p. 5163, Sep. 2020.

[23] R. P. Spang, W. Wardah, V. Schmitt, and S. Möller, "Unraveling the hangry rater: Non-linear effects of hunger on multimedia quality perception," in *Proc. 15th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2023, pp. 228–231.

[24] R. P. Spang, M. Warsinke, V. Schmitt, L.-F. Villa-Arenas, N. Ashrafi, and S. Möller, "Disentangling user states in QoE: Situation-dependent and independent factors," in *Proc. 16th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2024, pp. 132–138.

[25] D. J. McFarland and J. R. Wolpaw, "EEG-based brain–computer interfaces," *Current opinion Biomed. Eng.*, vol. 4, pp. 194–200, Jan. 2017.

[26] S. M. Hofmann, F. Klotzsche, A. Mariola, V. Nikulin, A. Villringer, and M. Gaebler, "Decoding subjective emotional arousal from EEG during an immersive virtual reality experience," *ELife*, vol. 10, p. 64812, Oct. 2021.

[27] C. Keighrey, R. Flynn, S. Murray, and N. Murray, "A physiology-based QoE comparison of interactive augmented reality, virtual reality and tablet-based applications," *IEEE Trans. Multimedia*, vol. 23, pp. 333–341, 2021.

[28] A. Zheleva, W. Durnez, K. Bombeke, G. Van Wallendael, and L. De Marez, "Seeing is believing: The effect of video quality on quality of experience in virtual reality," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–4.

[29] A. Zheleva, M. Kacper Gil, E. Pelosin, D. Talsma, L. De Marez, and K. Bombeke, "Kitchen horrors: Unraveling the influence of multimodal stressors on user experience in virtual reality through electrodermal activity," in *Proc. 15th Int. Conf. Appl. Human Factors Ergonom.*, vol. 137, 2024, pp. 10–19.

[30] B. Moharana, C. Keighrey, and N. Murray, "Physiological synchrony in a collaborative virtual reality task," in *Proc. 15th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2023, pp. 288–293.

[31] C. Cortés, I. Viola, J. Gutiérrez, J. Jansen, S. Subramanyam, E. Alexiou, P. Pérez, N. García, and P. César, "Delay threshold for social interaction in volumetric eXtended reality communication," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 20, no. 7, pp. 1–22, Jul. 2024.

[32] J. Sameri, S. Van Damme, S. Schwarzmann, Q. Wei, R. Trivisonno, F. De Turck, and M. T. Vega, "Collaborative cooking in VR: Effects of network distortion in multi-user virtual environments," in *Proc. ACM Multimedia Syst. Conf. ZZZ*, Apr. 2024, pp. 509–515.

[33] M. Nitzan and Z. Ovadia-Blechman, "Physical and physiological interpretations of the PPG signal," in *Photoplethysmography*, 2022, pp. 319–340.

[34] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," *Bioengineering*, vol. 3, no. 4, p. 21, Sep. 2016.

[35] P. van Gent, "Python heart rate analysis toolkit documentation," Tu Delft, Delft, The Netherlands, Tech. Rep., 2018. [Online]. Available: https://repository.tudelft.nl/record/uuid:5c638e14-d249-4116-aa05-2e566cf3df02

[36] Q. Xuan, J. Wu, J. Shen, X. Ji, Y. Lyu, and Y. Zhang, "Assessing cognitive load in adolescent and adult students using photoplethysmogram morphometrics," *Cognit. Neurodynamics*, vol. 14, no. 5, pp. 709–721, Oct. 2020.

[37] P. Celka, P. H. Charlton, B. Farukh, P. Chowienczyk, and J. Alastruey, "Influence of mental stress on the pulse wave features of photoplethysmograms," *Healthcare Technol. Lett.*, vol. 7, no. 1, pp. 7–12, Feb. 2020.

[38] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[39] M. Egger, M. Ley, and S. Hanke, "Emotion recognition from physiological signal analysis: A review," *Electron. Notes Theor. Comput. Sci.*, vol. 343, pp. 35–55, May 2019.

JAVAD SAMERI (Graduate Student Member, IEEE) received the M.Sc. degree in artificial intelligence and robotics from Iran University of Science and Technology, in 2019. He is currently a Doctoral Researcher with the Internet Technology and Data Science Laboratory (IDLab), imec, Ghent University, and the eMedia Research Laboratory, KU Leuven, Belgium. His research interests include physiological signal processing, human–machine interfaces, virtual reality (VR), and quality of experience (QoE) modeling.

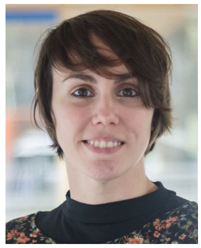
SAM VAN DAMME (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science engineering from Ghent University, Belgium, in 2016 and 2019, respectively, and the joint Ph.D. degree in computer science engineering and engineering technology from Ghent University and KU Leuven, Belgium, in 2024. He is currently a Postdoctoral Researcher with the Internet Technology and Data Science Laboratory (IDLab) Research Group, INTEC, imec, Ghent University. He has a strong interest and expertise in quality-of-experience (QoE) modeling and assessment, human–computer interaction (HCI), haptic feedback, and multi-user interaction with a main focus on immersive multimedia technologies, such as virtual and augmented reality (VR/AR).

SUSANNA SCHWARZMANN received the bachelor's and master's degrees in computer science from the University of Würzburg, Germany, and the Ph.D. degree from TU Berlin, in 2022. She was a Research Assistant with the Network Architectures Group. Since 2021, she has been with Huawei Technologies, Munich Research Center, where she holds the position of a Research Engineer with the Advanced Wireless Technologies Laboratory.

QING WEI received the B.Sc. degree in electrical engineering from Shanghai Jiao Tong University, China, in 1997, and the M.Sc. degree (Hons.) in communication engineering from TU Munich, Germany, in 2001. In 2002, she was with DOCOMO Euro Laboratories, leading the Mobile Optical Network Team. In 2015, she joined Huawei Technologies as a Principal Researcher and worked in the area of 5G mobile network architecture. Besides owning more than 50 publications and more than 100 IPRs in the area of mobility management and network slicing. She contributes directly to standardization activities, such as NGMN, 3GPP, and ETSI with more than 100 accepted TDoC submissions. She is the Task Leader in several EU projects, such as SASER-SaveNet, 5G-Xhaul, and 5G-MoNArch. Her research interests include 6G network architectures with embedded AI and enriched network services.

RICCARDO TRIVISONNO (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications engineering from the University of Bologna, Italy, in 2000 and 2005, respectively. In 2011, he joined Huawei Technologies, Germany, where he is currently the Head of Network Architecture—Research and Standardization—for the Advanced Wireless Technologies Laboratory, Munich Research Center. Over the past ten years, he has been a leading contributor to the definition and the standardization of 5G network architecture and technologies, delivering to 3GPP Releases 15–18—in the areas of architecture modularization, network slicing, network analytics, and QoS—filing more than 100 IPRs. He has been the Chairperson of 6G-IA Pre-Standardization WG, since 2020, and a Board Member of one6G Association since its foundation, in 2021. He was elected one6G Board Vice-Chair in 2023.

FILIP DE TURCK (Fellow, IEEE) leads the Network and Service Management Research Group, Ghent University, Belgium, and imec. He has involved in several research projects with industry and academia. He has co-authored over 700 peer-reviewed articles. His research interests include the design of secure and efficient softwarized networks and cloud systems. He has served as the Chair for the IEEE Technical Committee on Network Operations and Management (CNOM); and a Steering Committee Member for IFIP/IEEE IM, IEEE/IFIP NOMS, IEEE/IFIP CNSM, and IEEE NetSoft Conferences. In addition, he served as the Editor-in-Chief for IEEE Transactions on Network and Service Management.



MARIA TORRES VEGA (Senior Member, IEEE) received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Spain, in 2009, and the Ph.D. degree from Eindhoven University of Technology, The Netherlands, in 2017. She is currently a tenure-track Assistant Professor with KU Leuven, Belgium, where her research focuses on devising human-driven control and management mechanisms for enhancing the perception of immersive systems. Her research interests include quality of service and QoE in immersive multimedia systems and autonomous networks management.