UNICO I+D Project
6G-SORUS-RIS

SORUS-RIS-A2.2-E1

Clasificación inicial de UEs según el modelo de comportamiento

Abstract

This Deliverable describes the activities carried out during the first period of SORUS-RIS-A2.2. These include the identification of the psychological and behavioural concepts of relevance for the characterization of users' response to the system latency, as well as the identification of the methodology to be used for its measurement. Furthermore, this Deliverable presents the first version of the user classification algorithm, applied to three different datasets, as well as the results of the pilot study based on the use of psychophysiological signals, as a first step for the integration of these in the classification algorithm.

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES

Plan de Recuperación, Transformación y Resiliencia

Document properties

Document number	SORUS-RIS-A2.2-E1
Document title	SORUS-RIS-A2.2-E1. Clasificación inicial de UEs según el modelo de comportamiento
Document responsible	Ioannis Arapakis, Miguel Barreda
Document editor	Miguel Barreda
Editorial team	Ioannis Arapakis, Miguel Barreda
Target dissemination level	Public
Status of the document	Final
Version	1.0
Delivery date	19-12-2022
Actual delivery date	19-12-2022

Production properties

Reviewers	Ioannis Arapakis

Disclaimer

This document has been produced in the context of the SORUS-RIS Project. The research leading to these results has received funding from the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union-NextGenerationEU through the UNICO 5G I+D programme.

Content

Financiado por la Unión Europea NextGenerationEU

GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES

Plan de Recuperación, Transformación y Resiliencia

UNICO I+D

List of Figures

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

List of Tables

List of Acronyms

ECG - Electrocardiography

EDA – Electrodermal Activity

EEG – Electroencephalogram

EMG – Electromyography

ERP – Event-related potentials

FAA – Frontal Alpha Asymmetry

fNIRS – Functional Near Infra-Red Spectroscopy

HR – Heart rate

HRV – Heart rate variability

LPA – Latent Profile Analysis

PPG - Photoplethysmography

PSD - Power spectral density

QoE – Quality of Experience

SCL – Skin Conductance Level

SCR – Skin Conductance Response

Resumen Ejecutivo

En este Entregable se presentan las actividades realizadas durante el primer período de SORUS-RIS-A2.2. Estas incluyen las Tareas 1 y 2 definidas en la Oferta Técnica, así como la integración de los resultados de estas en una primera versión del algoritmo para la clasificación de los perfiles de usuario en función de su respuesta a la latencia.

La primera parte de estas actividades se ha basado en una revisión de literatura científica multidisciplinar, que ha permitido identificar los conceptos psicológicos clave respecto a la percepción de la latencia y su impacto en los usuarios (Tarea 1) así como las diferentes opciones metodológicas para medir este (Tarea 2). Los resultados de estas actividades han puesto de manifiesto la necesidad de abordar el impacto de la latencia tanto en la percepción subjetiva de la calidad de la experiencia de usuario como en varias dimensiones cognitivas y emocionales de la respuesta del usuario (atención, motivación, respuesta emocional, y engagement). También ha revelado la necesidad de una aproximación multimétodo a esta cuestión, que combine los métodos auto-informados con métricas obtenidas a partir de medidas psicofisiológicas. Las más adecuadas entre estas incluyen tanto métricas basadas en la señal de electroencefalograma (FAA, energía de la banda alfa occipital, energía de la banda beta parietal, etc.) como otras basadas en señales psicofisiológicas periféricas (EDA, medidas cardíacas, EMG).

Partiendo de esta caracterización, se ha llevado a cabo una doble estrategia para el perfilado de usuarios. En primer lugar, se ha diseñado un proceso en dos pasos para caracterizar la sensibilidad individual de los usuarios a la latencia y agruparlos en función de esta. Se han empleado tres datasets existentes disponibles online en los que se ha implementado este algoritmo de clasificación, permitiendo ejecutar el algoritmo y validar y reproducir la clasificación con varios conjuntos de datos diferentes. Los resultados han demostrado que, si bien la estructura básica de la clasificación permanece relativamente estable en diferentes conjuntos de datos, el contexto y las especificidades de estos también derivan en cambios importantes en los resultados, por lo que siempre deben ser tomados en consideración.

En segundo lugar, para integrar medidas psicofisiológicas y obtener una imagen más detallada del impacto de la latencia en los usuarios, se han implementado dos paradigmas experimentales: uno relativo al consumo de vídeos en streaming, y el otro relacionado con el uso de motores de búsqueda online. Se ha realizado un estudio piloto con el primero de ellos, y los resultados han permitido analizar el impacto de la latencia en diversas métricas psicofisiológicas, demostrando la utilidad de estas como medidas complementarias a los métodos auto-informados, proporcionando información útil para el empleo de estas de forma complementaria en la clasificación de perfiles de usuario que se llevará a cabo en la siguiente fase del proceso.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Executive Summary

This Deliverable presents the activities carried out during the first period of SORUS-RIS-A2.2. These include Tasks 1 and 2 defined in the 'Oferta Técnica', as well as the integration of the results of these tasks in a first version of the algorithm for the classification of user profiles according to their response to latency.

The first part of these activities has been based on a multidisciplinary scientific literature review, which has allowed identifying the key psychological concepts regarding the perception of latency and its impact on users (Task 1), as well as the different methodological options to measure it (Task 2). The results of these activities have highlighted the need to address the impact of latency, both on the subjective perception of the quality of the user experience as well as on various cognitive and emotional dimensions of user response (attention, motivation, emotional response, and engagement). It has also revealed the need for a multi-method approach to addressing this question, combining self-reported methods with metrics obtained from psychophysiological measures. The most suitable of these include both EEG signal-based metrics (FAA, occipital alpha band energy, parietal beta band energy, etc.) and others based on peripheral psychophysiological signals (EDA, cardiac measures, EMG).

Based on this characterization, a two-fold strategy for user profiling has been carried out. First, a two-step process has been designed to characterize the individual sensitivity of users to latency and to group them according to latency. Three existing datasets available online were used to implement this classification algorithm, allowing the algorithm to be run and the classification to be validated and reproduced with different datasets. The results have shown that while the basic structure of the classification remains relatively stable across different datasets, the context and specificities of the datasets also lead to important changes in the results and should always be taken into consideration.

Secondly, in order to integrate psychophysiological measures and obtain a more detailed picture of the impact of latency on users, two experimental paradigms have been implemented: one related to the consumption of streaming videos and another related to the use of online search engines. A pilot study has been carried out with the first one and the results have allowed to analyse the impact of latency on several psychophysiological metrics, demonstrating the usefulness of these as complementary measures to the self-reported methods and providing useful information for the use of these in a complementary way in the classification of user profiles that will be carried out in the next phase of the process.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

1. Introduction

The overall objective of the activity SORUS-RIS-A2.2 "Modelo de comportamiento del UE" is to identify user profiles that allow classifying users according to their cognitive, emotional, and behavioral response to the technical performance of the network. In particular, the focus is on the user's response to **latency** (e.g., discriminating between users with higher or lower latency tolerance). The rationale behind this objective is as follows: if information on these aspects (e.g., different tolerance to latency among users) is available, energy-sharing plans can be created that take advantage of the observed variability in latency perception and response to reduce processing costs (e.g., energy) when possible, without degrading the subjective experience of the users (e.g., prioritizing those users more sensitive to latency over those more tolerant to latency in a certain type of activity).

The document "Oferta técnica y plan de ejecución para "Caracterización de usuarios y emulación de escenarios 6G con superficies inteligentes reconfigurables en el marco del Plan de Recuperación, Transformación y Resiliencia – financiado por la Unión Europea –NextGenerationEU"" sets out the tasks required to progress towards this objective. In particular, for the project's first year, the tasks "Task 1. Definition and summarization of relevant user behaviours" and "Task 2. Definition and summarization of digital behavior sensors and channels" are included. These tasks involve the conceptual delimitation of the variables to be considered in user profiling, the methodologies to be used, as well as the identification of the best strategies in this respect.

This Deliverable describes the work carried out within the framework of these two tasks and their integration in a first approach to user profiling based on their response to system latency. Specifically, the work related to Task 1 is presented in section *2. Definition, identification, and summarization of relevant psychological constructs and aspects of user response and behavior*; whereas the work related to Task 2 is presented in section *3. Identification of relevant metrics, digital behaviour sensors, and channels*. The user profiling strategies implemented are presented in section *4. Strategies employed for user profiling* and, finally, section *5. Conclusions* summarizes the central insights obtained from the work conducted in this period.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

2. Definition, identification, and summarization of relevant psychological constructs and aspects of user response and behavior.

This section includes the work done on "Task 1. Definition and summarization of relevant user behaviours". This task is the first logical step in defining user profiles based on their response to latency: understanding to what we refer to when we talk about user response to latency. As described in the technical offer, this task aims to identify the relevant psychological constructs related to the perception of latency and the related cognitive, emotional, and behavioural aspects of users, which can then be mapped to indicators and signals (as described in Task 2). Such identification task is essentially a **conceptual analysis**, so the methodology employed has been based primarily on an extensive **review of the existing literature**. Specifically, an interdisciplinary literature review has been carried out, covering different interrelated research areas including, among others, research on quality of experience (QoE) carried out from disciplines related to multimedia engineering, research in the field of Human-Computer Interaction, research with a user experience (UX) perspective, as well as research from the perspective of media psychology and social communication.

The literature review and conceptual analysis carried out have allowed us to identify two main currents or singular approaches to the issue of the impact of latency on users. On the one hand, (i) we find the approach based on the **quality of experience**, focused on understanding the impact of the system's technical characteristics on the **user's subjective perception and conscious assessment**. On the other hand, (ii) we find the approaches that, instead of focusing on the user's subjective perception, analyze the **overall impact of the system's features on other descriptive variables** of the user's experience (such as, for example, their motivation for content consumption, the level of engagement with the content or reported enjoyment, etc.). The main aspects of these two approaches are described below.

2.1 Identification of relevant psychological constructs related to the subjective perception of latency and quality of experience

Latency in interactive systems is often defined as the time interval between an action performed by the user and its observable effect on the system (Halbhuber et al., 2023). It is composed of different elements, including input latency (the time between the user's input and its reception by the system), processing latency, which includes network latency and is central to this project, and output latency, which refers to the time between the completion of processing and the presentation of the output to the user (Wimmer et al. 2019). This definition of latency is particularly well suited to those interactive systems in which the user provides frequent inputs (such as search engines or other web applications, video games, etc.). However, even in those applications with a less interactive character, where user input is less frequent (e.g., video streaming applications), the same technical characteristics of the network that determine aspects of latency in processing contribute to determining the execution of the application and the subsequent user experience. Therefore, in this project, we will address different aspects related to system performance in a broad sense, not only related to latency in interactive systems but also to the impact of the aspects underlying latency in non-interactive systems (e.g., video streaming).

The literature review carried out for the identification of the relevant **psychological constructs** has highlighted the distinction between two key aspects of user experience: on the one hand, the ability to perceive latency in a system, and on the other hand, the impact of latency on the user's subjective assessment of the experience (Quality of Experience, QoE). Although both aspects are closely related, it is important to highlight the difference between them, since not all perceivable latencies necessarily impact the experience (e.g., Kaaresoja et al., 2014).

2.1.1 Latency perception

Latency perception refers to the user's ability to discriminate different levels of latency in a given system or interactive application. A key concept in this area is the "point of subjective simultaneity (PSS)", which comes from the field of cognitive and experimental psychology and can be understood as the level of asynchrony between two stimuli (usually expressed in milliseconds) at which they appear to be simultaneous to an observer (cf. Stone et al. 2001). Based on this concept, numerous investigations have explored what are the perceptual thresholds or PSS in different tasks or applications, using comparative methodologies that fall within the Just Noticiable Differences methodologies family.

The literature review on research based on this approach yields a number of key insights that should be taken into consideration in our project, namely:

- Latency perception thresholds are not something fixed and immutable but depend on various factors and combinations among them. Beyond technological factors, factors related to the **characteristics of the user** and the **task** to be performed play an important role.
- Among the characteristics of individuals, **age** and **experience** with the system seem to be key personal factors: the most experienced and the youngest perceive smaller latencies (Attig et al., 2017; Forch et al., 2017).
- With respect to task characteristics, small latencies are more likely to be perceived as annoying to the user when they occur in interactions that are short per se (e.g., Doherty & Sorenson, 2015)

Research along these lines has also provided different estimates of users' ability to perceive latencies in different contexts. While many guidelines and recommendations recommend a maximum latency of 100 ms for optimal user experience in basic interactions, the results of empirical work suggest that users can perceive much lower latencies. Some work shows that user performance

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

in zero-order tasks (and more demanding second-order tasks) can already suffer a clear deterioration with latencies between 16-60 ms (Attig et al., 2017). On the other hand, some work indicates that users can perceive very small latencies in some cases, even 1 or 2 ms (Ng et al., 2012, 2014). However, a key aspect that becomes apparent in the literature reviewed is the important distinction between the ability to perceive system latency consciously, and whether such latency impacts the user's subjective experience (e.g., Kaaresoja et al., 2014). Therefore, it is critical to analyze not only whether users perceive system latency, but also whether and how they report a possible impact of it on their user experience, as we will see below.

2.1.2 Subjective Effects of Latency: Quality of Experience

In addressing the subjective effects of latency, a relatively common perspective has been to focus on users' "tolerance to latency" (e.g., De Silva et al., 2010). From this perspective, user experience has been described in terms of acceptance or rejection of the experience, and the contribution of latency to that response has been analyzed. Such tolerance depends, on the one hand, on the perceptual limits of the user (i.e., whether or not he/she perceives a certain latency), but also on the user's expectations of the specific application or service, and the perceived complexity of the task the user is performing (e.g., Doherty & Sorenson, 2015).

A more narrowed-down view than the one premised on latency tolerance is the approach adopted by research on Quality **of Experience (QoE)**. More specifically, QoE has been defined as "the degree of delight or annoyance of the user of an application or service" (Le Callet et al., 2013). Although the use of terminology is not uniform among researchers working in this field, the concept of QoE differs from that of quality of service (Quality of Service, QoS: "the totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service; ITU, 2019) in that the former is much broader, and is influenced by human and contextual factors that go beyond the technical characteristics of the system usually encompassed within the concept of QoS (latency, jitter, packet loss, etc.) (Varela et al., 2014).

Previous research has identified **three types** of factors that have an impact on QoE: (1) **human factors**, which include, among others, aspects such as the person's visual or hearing acuity, or socio-demographic or cultural aspects; (2) **system factors**, which include those related to the type of content, those related to the medium used (e.g., resolution, frame rate, etc.), those related to the network and transmission, and the device used, etc.; and, finally (3) **contextual factors** (e.g., task type, physical, temporal, social, economic, etc.) (Reiter et al., 2014; ITU, 2019, 2021). The numerous factors that influence QoE (and the fact that many of these are intrinsically interrelated) illustrate the complexity of QoE assessment and highlight (i) the need to approach QoE based on the specific characteristics of the technology to be assessed, and (ii) the need to take into consideration the context of specific use cases and the particularities of users.

In regards to the types of technology, much of the research on QoE, originating mainly from the field of multimedia and computer engineering, has focused on the **perceptual aspects** of

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

multimedia systems (in terms, for example, of perceptual image or audio quality) (Raake & Egger, 2014). To this end, subjective methodologies have often been employed in which users are asked to evaluate the quality of the video, image, or audio, depending on the type of stimulus and the objectives of the study and report this on a scale. The scales used for such analyses depend on the type of methodology employed: for example, in **single stimulus methodology**, participants report perceived quality on a scale ranging from *Poor* to *Excellent*, whereas if methods involve the comparison of two stimuli, comparative scales (e.g., ranging from *Much Worse* to *Much Better*) may be used (ITU, 2019). Although this approach has been commonly used to evaluate image quality in video content, the concept of QoE has also been applied in more interactive applications like video games, where latency can play a central role. For example, research (Liu et al. 2021) suggests that latencies around 150 ms considerably impact the reported QoE, implying a reduction of up to 25%. On the other hand, regarding users' characteristics, there is evidence that cultural and personality traits play a central role in QoE. Specifically, the study conducted by Scott et al. (2015) shows that **personality and cultural traits** represent 9.3% of the variance attributable to human factors and, more importantly, that human factors explain an equal or higher proportion of variance compared to technical factors. Personality factors have been shown to be relevant for quality estimations in several contexts, including visual and audio QoE (Galloso et al., 2016; Scott et al., 2015; Wechsung et al., 2011).

2.2 Identification of relevant user behaviours and cognitive and emotional responses

Beyond the user's ability to detect latency or the impact of latency on the overall assessment of the quality of the user experience, numerous research studies have addressed a different approach: analysing the impact of technical qualities of systems (such as latency) on **cognitive or emotional processes,** or specific **user behaviours** in various contexts of use. These approaches are fundamental to understanding user responses, as research shows that the perception of quality and its impact on other key aspects of the user experience are not always closely associated. For example, whether an experience is perceived as more or less quality does not always imply that it is perceived as more or less enjoyable: factors such as the **enjoyment of content** may be more determined by non-technical aspects (such as the presence or not of co-viewers, e.g., Zhu et al. 2015) than by the technical aspects of the content.

Therefore, in addition to perceptual quality assessments, such as those mentioned above, when evaluating the QoE of a system, it is also essential to consider theoretical and methodological approaches from other research fields, such as theoretical models on user acceptance of technology (e.g., Technology Acceptance Model, TAM) and User Experience (UX) research. In the technology acceptance research, the focus has been on psychological constructs such as the user's **attitude** (positive or negative) towards technology and intention to use it, as well as variables that can predict these, such as the ease of use of such technology or perceived usefulness (Marangunić & Granić, 2015). On the other hand, UX research, while largely sharing the objectives of QoE research, is often characterized by addressing aspects of user experience that go beyond the purely perceptual (e.g., user needs, experienced emotions), as well as by a greater prominence of qualitative methods (e.g., semi-structured interviews) (Wechsung & De Moor, 2014). In particular, this methodological approach, can be of great interest in revealing aspects of the experience not initially considered, which is fundamental in research with novel technologies.

A construct that has been considered key in numerous research studies in this regard is the idea of **engagement** with a certain technology, application, or service. It has been defined in various ways (cf. O'Brien, 2016): for example, engagement has been conceptualized as "the state of mind that we must maintain in order to enjoy a representation of an action" (Laurel, 1993), in a sense related to attention and cognitive resource allocation. More closely related to the idea of flow, engagement has also been defined as a component of a system's usability that "encourages interactions" (Quesenbery, 2014) or evokes a "state of playfulness" (Webster & Ho, 1997), or as a "a user's response to an interaction that gains, maintains, and encourages their attention, particularly when they are intrinsically motivated" (Jacques, 1996). A review of the different subcomponents of engagement present across different theories and authors shows that many of the approaches to this topic recognize several key subcomponents of engagement that include attention, motivation, and user emotional response as key factors (O'Brien & Toms, 2008; Lalmas et al., 2022).

Complementarily, some authors have considered that an essential part of engagement goes beyond psychological dimensions and involves user **behavior**, for example in terms of continued use of a certain technology over time, or in terms of performance with the technology (e.g., the performance of a video game player). In that respect, some prior research has explored the impact of latency on some of these variables. For instance, Barreda-Ángeles et al. (2015) examined how search engine latency impacts user attention, emotional response, and subsequent usage behavior. In the context of videogames, Halbhuber et al. (2021) showed that reducing latency in an FPS game from 180 ms to 60 ms increases the positive affect associated with the gaming session and that latency harms enjoyment and performance. Regarding the performance of the video game user, some studies (Annett et al. 2014; Jota et al. 2013) suggest that performance deteriorates when latency exceeds 25 ms, but no improvements were observed for latencies below 25 ms. Thus, values of 25 ms seem to represent or below seem to represent the optimal latency in this context. Other work by Durnez et al. (2021) examined the effects of latency on game experience in a desktop-based exergame and found that latency decreases the experienced flow, with flow referring to a state of complete immersion, effortless concentration, and enjoyment.

We note that other works have opted for different concepts that aim to provide a snapshot of the state of the user as a function of the technology employed (e.g., the concept of frustration; Long et al., 2018). In this sense, the particularities of the technology with clear determinants of which concepts are most critical when evaluating user experience. Such an example are the concepts used to evaluate user experience in virtual reality. Specifically, Debarba et al. (2022) evaluate the quality of experience with 3D rendered virtual reality environments and stereoscopic 360 video using three factors: the **user's perception of the quality** of the content, whether the user experiences

cybersickness, and the reported **feeling of presence** ("being there"). Singh et al. (2022), on the other hand, evaluate the quality of the experience in a VR environment for business meetings by focusing on factors, such as the feeling of immersion and social presence ("being there with others") reported by participants, as well as the scores given by participants to the system in terms of usability, embodiment, quality of interaction, and quality of communication. Despite the differences in their conceptualization of QoE and its component aspects, it is clear from these studies that the evaluation of feelings of presence in immersive environments is a key aspect in describing the user experience. Also, in the same vein, and to provide a unified approach to the evaluation of QoE in immersive communication systems, Toet et al. (2022) propose a holistic approach in which QoE in immersive communication systems that evaluates feelings of spatial and social presence in five dimensions, which refer to the sensory, emotional, cognitive, reasoning, and behavioural qualities of the user during interaction with the system, and with other users through it.

Thus, while some more generic concepts such as **engagement** (and its subcomponents in terms of **attention**, **motivation**, and **emotional response**) seem to have relevance in most contexts and use cases, other concepts such as **enjoyment**, **immersion**, or **presence**, may be particularly relevant in specific use cases.

3. Identification of relevant metrics, digital behaviour sensors, and channels

This section addresses "Task 2. Definition and summarization of digital behavior sensors and channels". The work done in this task is based on a **review of the existing literature** for the identification of metrics and methodologies for the measurement of psychological and behavioural constructs, as well as on the analysis of their usefulness in different hypothetical use cases.

The most common approach to user experience measurement is based on the use of questionnaires (e.g., ITU, 2019). However, it is well known that humans do not have conscious access to all our cognitive and emotional processes, which can determine our behaviour (Nisbett & Wilson, 1977). Therefore, individual and personalized measurement of various aspects of user experience through the use of psychophysiological measures is a topic that is gaining more and more weight in the study of user experience (Yamazaki, 2021) and, in general, there is consensus that psychophysiological approaches are necessary for a more complete view of user experience beyond the simple use of questionnaires (Engelke et al., 2016). Thus, in this section we consider two dimensions: (1) Using self-reported measures in the format of questionnaires and experience sampling methods, which can have great utility in capturing user experience in real environments, and (2) Using (objective) psychophysiological measures, which can provide reliable and objective indicators of the psychological processes described in the previous section, while avoiding the direct question to the user (self-report).

3.1 Questionnaires and Experience Sampling

The use of **questionnaires** is the most traditional methodological approach in QoE research and other psychological aspects of user experience (Díaz-Oreiro et al., 2019; ITU, 2019). In addition to their proven psychometric properties, the use of validated questionnaires is relatively low cost and relatively easy to administer and analyse. However, the temporal resolution that questionnaires can offer is often relatively poor, and in many cases (for example, when they are not administered immediately after the experience to be evaluated) they rely on the user's ability to recall it, which can reduce their reliability. This problem can be solved for the most part with the use of experience sampling.

The **experience sampling method** (Csikszentmihalyi et al., 2014; Fischer, 2009; Van Berkel et al., 2017) asks the participants to report on aspects of their experience at various times of the day. In its modern form, this method may involve the use of an app on the person's cell phone, which triggers and reminders at certain times. In this way, it bears similarities to the use of diaries, but has some advantages over it, including:

- By asking the participant to evaluate their experience **immediately** after it occurs, it ensures that they remember it well and do not forget key aspects.

- The use of the cell phone allows the **parallel collection of other data** (e.g., about the apps the person was using at the time, etc.), which is extremely useful for studies on the quality of the experience in this context.
- The use of the cell phone also allows the presentation of questionnaires with **complex logics**, if necessary, and the presentation of complementary media (e.g., photographs or graphs, which may be of interest in some types of studies).

However, the experience sampling method also has some important limitations, such as those mentioned below:

- The **burden** for the participant, as well as the interruption of daily activities, can lead to a distortion in the participant's experience that can affect the assessments and the quality of the data collected.
- As a consequence of the above, the level of participant **attrition** in studies using this method is traditionally high.
- This method often requires the installation of an app on the participant's cell phone. In addition to the cost of this app, this may have implications on the perception of issues such as **privacy**, etc. by the participants.

Finally, a very relevant aspect of these is that, despite its advantages, the experience sampling method is still a self-reported method, so it has some of the basic limitations of questionnaires: it depends on conscious access to psychological and emotional processes by the users and is subject to their cognitive biases.

3.2 Psychophysiological Measures

Psychophysiological measurements can be divided into two main categories: (1) Measurements of the central nervous system, and (2) Measurements of the peripheral nervous system.

3.2.1 Measurement of central nervous system activity

The main methods for measuring the activity of the central nervous system in the context of QoE and user experience research are electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS).

3.2.1.1 Electroencephalography

Electroencephalography has been one of the most powerful techniques for studying brain activity in a non-invasive manner. It involves placing a set of electrodes on the participant's scalp, to measure the electrical activity that arises from various regions in the cerebral cortex (Baillet, 2017). Through EEG recordings (Figure 1), one can infer a person’s cognitive state (e.g., mental workload) or simple thoughts such as moving an arm (motor imagery) on a real-time basis. Due to its high temporal resolution, low cost and portability, EEG can be utilised for brain-computer interface (BCI) systems.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

The scalp EEG signal has a variable amplitude that depends on many factors, such as origin and the number of brain sources, electrode montage (grand average, bipolar, etc), sensor quality (gel-based vs dry-based), etc. It is fair to say that most of the normative EEG amplitudes range around 10-100 µV. Typically, EEG systems acquire data with a bit depth of at least 16 bits per sample and a minimum sampling rate of between 256 Hz up to 5 KHz. For non-invasive EEG measurements, the brain's most reliable frequency bandwidth goes between 0.1-20 Hz. The number of electrodes found in EEG systems varies from 8 up to 256 channels. The electrode placement for each system is often a customised or extended version of the international 10-20 system.

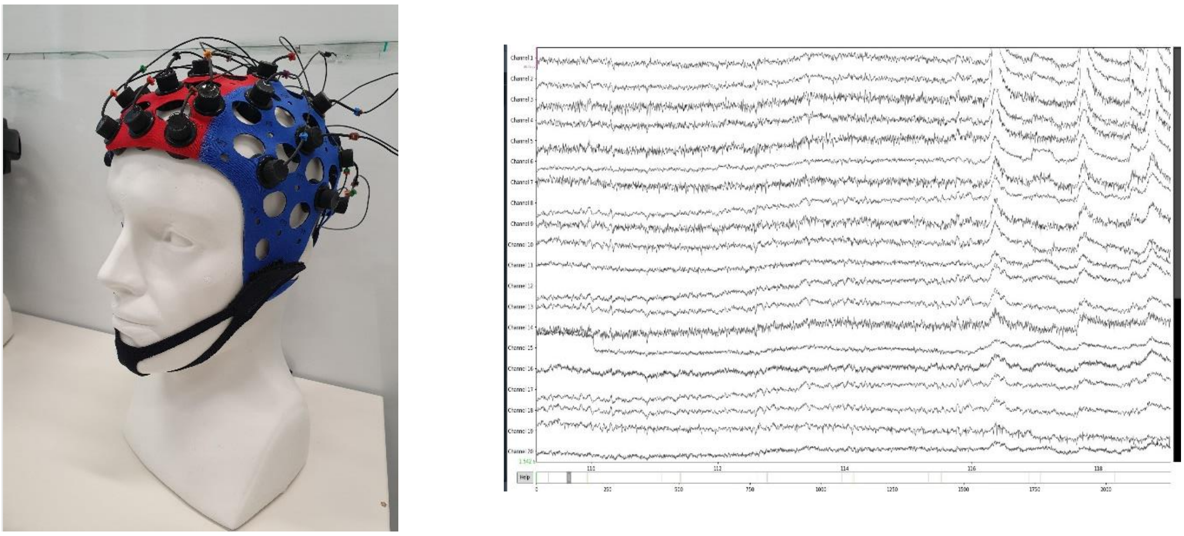

FIGURE 1. CAP FOR THE COLLECTION OF EEG AND SIGNALS SAMPLE

The most widely used strategy in the literature for interpreting EEG data has been the **spectrum band approach** which, interestingly, is also the first one utilised in early EEG studies. Specific EEG frequency bands have been associated with sleep stages and levels of alertness, and are commonly defined as follows (Martinek et al., 2021):

- 1-4 Hz: Delta band – found in the deep stages of sleep.
- 4-8 Hz: Theta band – found in the initial stages of sleep.
- 8-12 Hz: Alpha band – found at increased levels during wakeful relaxation, and particularly when the person has his/her eyes closed.
- 12-30 Hz: Beta band – found when the person is alerted or is engaged in mental activity.
- 30-80 Hz: Gamma band – observed when the person is doing a high cognitive task such as a working memory or reading task.

A sliding-window approach is employed when computing the power spectral density (PSD) in EEG to track variations in the power of each frequency band. Typically, a 30 sec time window is

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

considered, although shorter time windows (e.g., 3 sec) are also sometimes preferred for achieving higher temporal resolution (Olbrich et al., 2009).

Another approach for detecting certain psychological states is the analysis of **event-related potentials (ERPs)**. It refers to average responses time-locked to the onset of a series of identical or very similar stimuli. Through the process of averaging, the background noise is suppressed allowing consistent patterns of EEG activity to be revealed. ERPs consist of positive and negative deflections in the voltage which can be studied based on the amplitude and latency of each peak. The peaks are labelled based on their polarity (positive or negative) and order (e.g., P1, N1, P2, N2) or approximate time in milliseconds (e.g., N100, P200, P300). However, it is important to notice that the use of ERP analysis in QoE and user experience research has been somewhat limited as, by definition, ERP analysis requires the presence of discrete events or stimuli and, thus, it cannot be applied in studies where participants are engaged in an ongoing task (e.g., using a search engine, playing a video game or watching a YouTube video). Thus, although some studies have tried to use ERPs for the analysis of quality of experience (Arndt et al., 2014; Scholer et al., 2012), most studies in this area employ the frequency bands approach.

Several investigations have allowed defining different metrics, based on the frequency band approach, to obtain real-time EEG measures of some of the key constructs to characterize the user experience. These are summarized in Table 1.

Table 1. Metrics for psychological constructs based on EEG

Construct	Metric	Description	Refs
Motivation (motivational approach)	Frontal alpha asymmetry	The difference in log-transformed power of the alpha band between left and right frontal sites (ln[right]-ln[left] alpha power)	Arapakis et al., 2017; Kroupi et al., 2014; Smith et al., 2017
Visual attention	Occipital alpha	Power of alpha band in occipital sites	Smith & Gevins, 2004; Thut et al., 2006
Negative emotional re-action	Parietal beta	Power of beta band in parietal sites	Tao et al., 2019
Engagement	Engagement in-dex - frontal	Power of beta band divided by the sum of the power of theta and alpha bands (all in frontal sites)	Szafir & Mutlu, 2013
	Engagement in-dex - parietal	Power of beta band divided by the sum of the power of theta and alpha bands (all in parietal sites)	Nuamah & Seong, 2018

3.2.1.2 Functional near-infrared spectrography

Functional near-infrared spectroscopy is a non-invasive optical imaging technique for studying brain function (Boas et al., 2014). It relies on the fact that near-infrared (NIR) light can travel through the human scalp and skull, and reach the underlying neuronal tissues in the cerebral cortex. The amount of backscattered light is captured by a detector, which is used to measure changes in oxygenated (HbO2) and deoxygenated (HbR) haemoglobin following neuronal activation.

fNIRS systems are becoming more and more popular over the last decade since they are less prone to motion artifacts as compared to other neuroimaging modalities (e.g., EEG) and some systems also offer portability as they are small enough to be wearable and wireless (Perrey, 2008). As such, fNIRS researchers have recently started exploring the potential of this technology for studying brain function in more naturalistic settings than the lab.

With respect to the required instrumentation for fNIRS measurement, a light detector is placed at a certain distance from the NIR light source to collect the backscattered light and measure changes in light attenuation. The measured light can be used to infer brain activity from a "banana-shaped" brain volume along the path between the source and the detector. Features obtained with fNIRS related to the regional activity (e.g., moving average of HbO2 levels) and connectivity between pairs of regions (wavelet coherence) have been shown to greatly benefit Machine Learning (ML) and DL approaches in detecting mental states (e.g., drowsiness) (e.g., Khan et al., 2019).

However, since it is a much more recent method than EEG, and perhaps also because of the greater complexity in its use and analysis, the use of fNIRS in QoE and user experience research is still minimal.

3.2.2 Measurement of peripheral nervous system activity

Within this group of measures, the ones with the highest relevance in the domain of QoE and user experience measurement are measures **of electrodermal activity**, **cardiac activity**, and **facial electromyography**.

3.2.2.1 Electrodermal activity

Electrodermal activity (EDA) refers to changes in skin conductivity due to sympathetic nervous system activity. The activation of the sympathetic branch of the nervous system stimulates the production of sweat in the eccrine glands located in the palms of the hands and soles of the feet, increasing skin conductivity in these areas. Thus, by passing a small electric current between two electrodes placed in these areas (usually on the palmar side of two fingers) it is possible to measure such changes in conductivity, obtaining a proxy of sympathetic activation (Figure 2). The result is a non-stationary signal, in which two components can usually be distinguished: a tonic component - often referred to as **Skin Conductance Level (SCL),** which represents the overall level of skin conductance and varies relatively slowly, and a phasic component - **Skin Conductance Responses (SCR)** - which involves momentary, relatively faster, increases over the tidal drift of the SCL (Boucsein, 2012, Dawson et al., 2007). The study of EDA has been one of the most popular psychophysiological signals (Dawson et al., 2007). This is due, in part, to the relative simplicity and the low cost of the equipment needed to collect it but also to the fact that EDA can provide information about numerous mental constructs involving changes in sympathetic activity. As a proxy of sympathetic activity, EDA is considered "a pure arousal indicator" (Nardelli et al., 2022), which, in turn, underlies various cognitive and emotional processes.

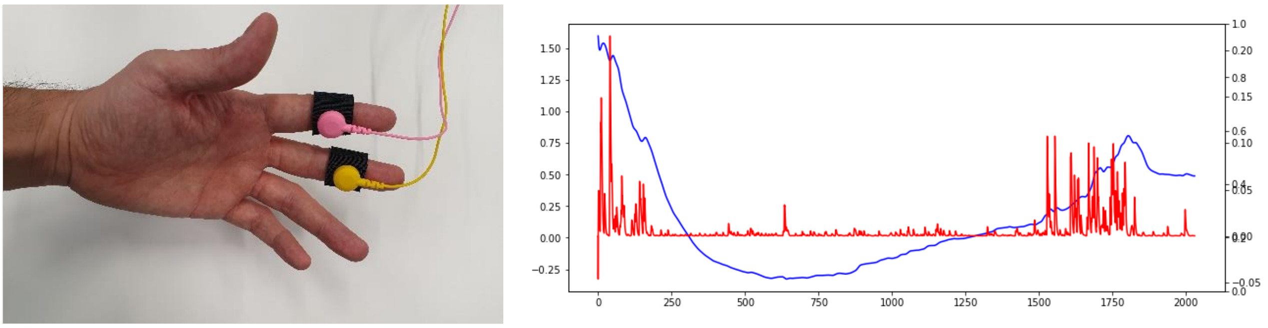

FIGURE 2. EDA SENSORS AND SIGNAL SAMPLE

3.2.2.2 Cardiac activity

Heart rate (HR) varies on a moment-to-moment basis in response to ongoing changes in physical and cognitive demands and is regulated by the two branches of the autonomic nervous system, namely the sympathetic and parasympathetic components (Rajendra Acharya et al., 2006). Increased activity of the sympathetic component is typically characterized by elevated HR and decreased **heart rate variability (HRV)**, while increased parasympathetic activity is characterized by decreased HR and increased HRV. Considerable evidence exists that the evaluation of the momentary changes in HRV can provide surrogates of fluctuations in cognitive processes, and, in the field of user experience and the measurement of psychological aspects of media, heart rate measurements are commonly used as indicators of attention and emotional arousal (Bolls et al., 2019).

Experiments in laboratory settings often employ **electrocardiography (ECG)** for monitoring cardiac activity which involves the placement of three or more lead cables on the upper chest and lower abdomen. ECG measures the electrical activity that arises from heart muscles during cardiac contractions and passes through the soft tissues to the superficial skin. The basic pattern of the ECG consists of a series of waves (deflections of electrical activity), including the R wave that reflects depolarisation of the main mass of the ventricles and, thus, it is the largest wave. As such, the HR is defined based on the R-to-R (RR) intervals across the cardiac cycles. The HR is also often measured in a laboratory with **photoplethysmography (PPG)** placed on the finger (Figure 3). PPG is an optical-based device that detects changes in blood volume and is used for determining the HR based on the peak-to-peak (PP) intervals of the acquired signal. Although PPG measures a hemodynamic signal rather than the electrical activity of the heart, it provides similar HRV traces to the ones obtained using ECG (Nardelli et al., 2020). In addition, cardiac activity is sometimes monitored with wearable

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

devices (e.g., smartwatches) that are well tolerated by participants and are more suitable for naturalistic experiments that involve movement.

Several HRV measures have been proposed in the literature which can be broadly categorised into **time-domain** and **frequency-domain** measures (Shaffer & Ginsberg, 2017). For time-domain measures, the fluctuations in instantaneous HR or the time intervals between adjacent heartbeats are first determined. In the case of frequency-domain HRV measures, the time series of beat-to-beat intervals is first derived and resampled to a timeline with equidistant intervals (e.g., 100 ms). Subsequently, the power spectral density (PSD) is estimated using non-parametric (e.g., fast Fourier transform) or parametric techniques (e.g., Welch’s method) that allows the separation of HRV into distinct frequency bands, related to different cognitive states.

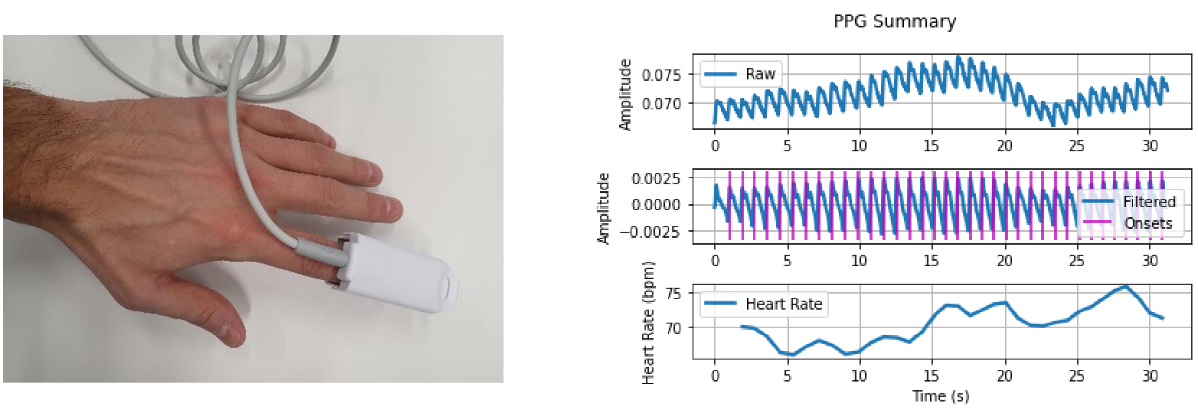

FIGURE 3. BLOOD VOLUME PRESSURE SENSORS AND SIGNAL SAMPLE

3.2.2.3 Facial electromyography

Facial electromyography (facial EMG) refers to the measurement of the electrical activity of certain muscles of the face, which can provide indications of an individual's **emotional states** (Mauss and Robinson, 2009; Wolf, 2015). In a dimensional model of emotions (Russell, 1980), emotion valence refers to whether the emotion is positive or negative, ranging from pleasant to unpleasant. The use of facial electromyography often focuses on the measurement of two facial muscles in particular: corrugator supercilii and zigomaticus major, which participate, respectively, in actions such as frowning and smiling. Activity in corrugator is linearly related to a negative valence of emotions and has been consistently used to distinguish between pleasant and unpleasant stimuli. In turn, activity in the zygomatic muscle is related to positive valence of emotions (Baur et al., 2015). Facial EMG is generally recorded bipolarly with small surface electrodes (contact area diameter ≤ 4 mm) located close to each other.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

4. Strategies employed for user profiling

Once the different relevant psychological and behavioural cues and the appropriate methods and signals for their measurement have been identified, the next step is to use them for the analysis of the different profiles existing among users, according to their response to latency. For this purpose, a two-step strategy has been followed, as described below:

1. First, we analysed the impact of latency on **generic subjective measures of quality of experience** (following the conceptual approaches described in section 2.1.2 and the methods described in section 3.1). By focusing on generic effects (subjective quality perception), this approach offers the advantage of exploiting available online databases, thus facilitating the profiling of large amounts of subjective experience data in real environments. This approach is described in section 4.1.
2. Second, the impact of latency on **various cognitive and emotional dimensions** of user experience was analysed (following the conceptual approach described in section 2.2), integrating the use of psychophysiological signals (as described in section 3.2). This allows a much more fine-grained understanding of the impact of latency on user experience, facilitating more specific profiling. This approach is described in section 4.2.

4.1 User profiling based on generic QoE measures

For this task, existing and available online datasets have been used, which allow access to a large amount of data on user responses to different latency conditions in real environments or in realistic simulations of real experiences.

4.1.1 Description of the dataset

The dataset **"QoE App Rating Dataset"** (Boz et al., 2019) was selected for this task. This dataset contains 64,036 observations (from a diverse group of 287 users in Finland). Participants installed custom software on their cell phones that collected network quality information (e.g., type of network, round trip time, etc.) and information from the apps that the users were employing. It also asked users to rate the quality of their experience on a scale of 1 to 5. The vast majority of observations (74%), though not all, are associated with Wi-Fi or LTE (4G) networks. The dataset also contains some socio-demographic data of the participants, such as their gender (Female: 143 participants; Male: 136; Other/NA: 9) or age group (16-25 years old: 133 participants; 26-40 years old: 129; +40: 25).

For each participant, the dataset contains between 50 and 2734 observations, with a mean of 223 observations per participant (SD = 249.8). Participants used between 1 and 26 applications in this period, with a mean of 9.8 applications per participant (SD = 3.66). This wealth of data per participant makes this dataset a valuable resource for user profiling. Regarding the applications used, most of the observations belong to apps from the *Social* category (23568 observations, 37% of the sample), followed by the *Communication* category (18877, 29%), followed by *Video Players and Editors* (4930, 8%), *News and Magazines* (3746, 6%), and *Music and Audio* (1915, 3%), in addition to other categories with a smaller number of observations.

The authors' analysis of this dataset allowed them to draw conclusions such as that the features that are important for predicting QoE in one type of application are not important for another type of application, or that beyond a certain limit, a large number of the events in which poor QoE is reported are related to non-network factors such as app quality, device performance, and user expectations, etc. Likewise, this analysis showed that, when predicting QoE, the most important feature is smartphone usage years (which, according to the authors, potentially acts as a proxy for technological sophistication, and which may influence user expectations). However, this dataset **has not been used to analyse the different individual responses** to network characteristics and, in particular, to latency.

4.1.2 Method

In order to establish examine whether there are groups or clusters of users based on their response to latency in the network, the method described below has been designed. It is an algorithm that performs two steps:

1. To establish **individual sensitivity** to latency (i.e., how latency affects the perception of the quality of the experience at the individual level). To this end, we explored for every participant the correlation between variations in latency levels in the different observations of that participant and the quality scores they provided. Given that other factors have a great weight in the quality assessment (for example, in our setting, the specific app that was being used, or the memory available on the participant's device), the influence of these factors was controlled. For this purpose, a **multiple regression** was performed, in which the quality assessment (normalized for each participant, as z-scores) was taken as output, and as predictors the values of 'round trip time' (normalized, z-scores, per participant) as a proxy for latency, as well as the available memory in the device and the app scores (in the Android App Store) that were being used at the time (to control the influence of these two factors). The resulting **coefficient** for 'round trip time' in the regression, as well as the **p-value** associated with it, were extracted as measures of the participant's sensitivity to latency and the variability of this measure (lower p-values represent measures with less variability), as a proxy for that participant's consistency in their sensitivity. This process has been implemented in Python.

2. The sensitivity data for each participant were then used to **estimate clusters of participants**. To this end, we initially explored the optimal number of clusters using the elbow method and a visual analysis, as well as the silhouette method. Once the optimal number of clusters was determined, a **k-means clustering** was performed. This was complemented by a **Latent Profile Analysis (LPA)** (Spurk et al., 2020). In this analysis, it is assumed that the variance within a population can be minimized by introducing a categorical latent variable. This latent

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

variable effectively divides the population into two or more subgroups that exhibit greater homogeneity in their patterns of variable means and covariances. This method adds some advantages to non-latent cluster methods (such as k-means), allowing a more nuanced understanding of individuals' cluster memberships (e.g., accommodating fractional or partial memberships) and models unobserved heterogeneity within the data, allowing for the identification of latent structures that traditional clustering methods might overlook (Magidson & Vermunt, 2002). To apply the LPA, starting from the number of clusters identified, we used the tidyLPA package (Rosenberg et al., 2019) in R, comparing solutions with equal" and "varying" variances and "varying", "equal", and "zero" covariances, and retaining the best solution according to the hierarchical approximation of Akogul & Erisoglu (2017).

4.1.3 Results

To increase the realism of the results taking into account the current characteristics of the technology, from the original dataset we retained only the observations that make include Wi-Fi and LTE networks, and those users who had more than one observation for these networks. This leaves a final sample of 48494 observations from 263 participants.

The analysis based on the elbow and silhouette method (Figure 4) seems to yield different results: while the elbow method seems to suggest that the optimal number of clusters is 3, the silhouette analysis indicates that the optimal number is two. Since a larger number of clusters can provide a more detailed understanding of the phenomenon, we chose to use three clusters in the k-means method.

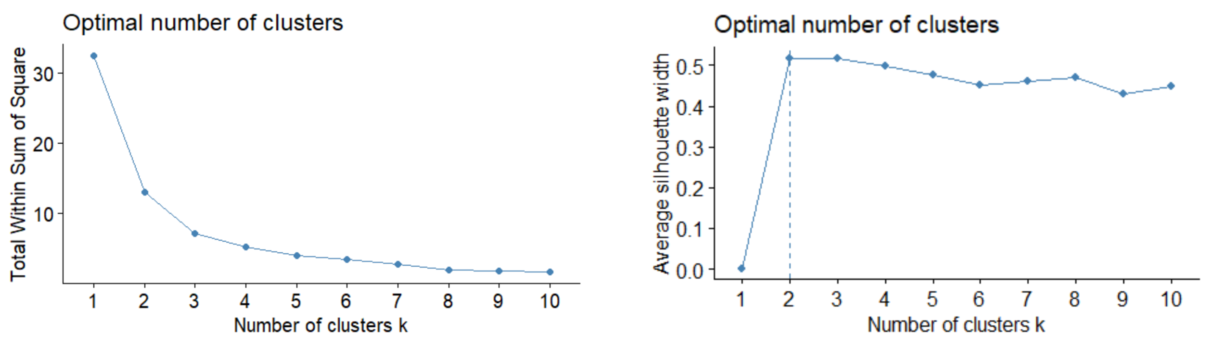

FIGURE 4. ELBOW AND SILHOUETTE GRAPHS FOR THE QOE APP RATING DATASET

The resulting data (Table 2) show the following cluster characteristics. Cluster 1, which includes 90 users (34% of the sample) to contain those users where there is no correlation between network latency and the quality they report (coefficient almost equal to 0, and very high p-value). In the second group, there is a correlation that appears positive, but is far from significant (average p-value considerably high). In this group, this positive correlation goes against logic (higher latency should be associated with lower quality) and is possibly an artifact because of the multiple regression from which this value is obtained (but which, as we said, seems not to be stable, based on the average p-value). This cluster contains 98 users (37% of the sample). Finally, a third cluster shows a coefficient with a much higher absolute value and an average p-value close to statistical significance (at $p < .05$). For participants in this cluster, latency is associated with a decrease in reported quality of experience: for a particular participant, a 1 standard deviation increase in latency experienced means a 0.24 standard deviation decrease in the quality score provided. This group of latency-sensitive participants contains 75 users (29% of the sample).

Table 2. Clusters obtained from the dataset "QoE App Rating Dataset" (Boz et al., 2019)

Cluster number	Variable means		N users in cluster
	"coefficient"	"p-value"	
1	0.01	0.76	90 (34%)
2	0.06	0.32	98 (37%)
3	-0.24	0.06	75 (29%)

Thus, there appear to be two clusters (1 and 2) of users in which there is no clear relationship between latency and quality of experience, and a third group in which there is. An LPA was then performed considering three clusters, as described above. This analysis shows that, indeed, the two groups of users with little or no sensitivity to latency (classes 2 and 3, in Figure 5), hardly differ in their average sensitivity to latency (variable coef_lat, in the Figure); rather, they only show a clear difference in the p-value associated with latency. This suggests that both groups are generally insensitive, although in one of the cases the user responses appear to vary more in their level of sensitivity. A single group of users (class 1, in the figure) has a clear sensitivity to latency, which consistently impacts their quality of experience.

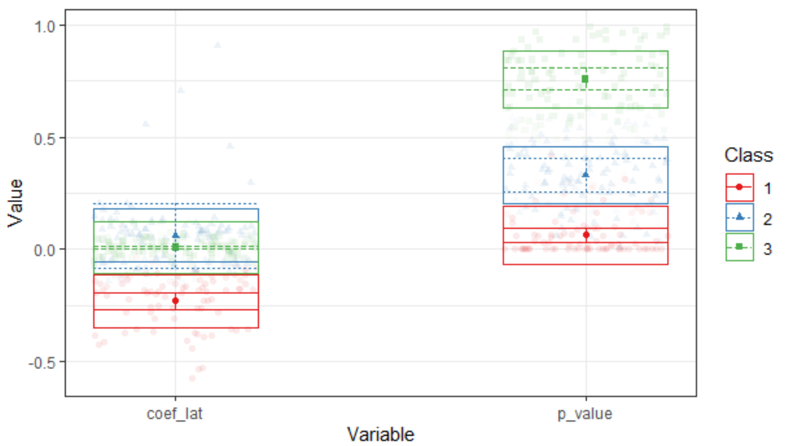

FIGURE 5. RESULTS OF THE LPA ANALYSIS CONDUCTED IN THE QOE APP RATING DATASET

4.1.4 Reproducibility of the classification with other datasets

Having established a first approximation to the definition of user profiles based on their response to latency, we then analysed whether these profiles are stable and reproducible with different datasets. A limitation found in this sense is that the available datasets present a high variability in the type of technology use situations that are collected in the dataset, as well as in the variables included. For this reason, it is necessary in some cases to use different proxies which, although they make comparison possible, prevent an exact equivalence in this comparison. For this task, two datasets have been used, one of which contains data related to a streaming video viewing context, while the other contains web browsing tasks.

4.1.4.1 Dataset on QoE in video streaming

To analyse the reproducibility of user profiles in the context of QoE evaluation in streaming video, the 'Pokemon' dataset (Amour et al., 2015) was used. It contains 1543 observations, from 181 users (of which 166 had valid data to be included in our analysis), who rated the quality of streaming videos on mobile devices, while different aspects of QoE were also recorded (more details can be found in the original article, Amour et al., 2015).

Since there are no round-trip time measurements in this dataset, we employed a proxy for system latency: the buffering time measurement, which is a variable that is also related to network capacity and is a useful proxy in a context of low interactive use, such as video viewing. As a measure of quality, we used the mean opinion score (MOS) of these videos, as reported by the participants. We calculated the correlation between these two variables (buffering time and MOS) for each participant, as a measure of their sensitivity to latency, and performed the clustering procedure as described above.

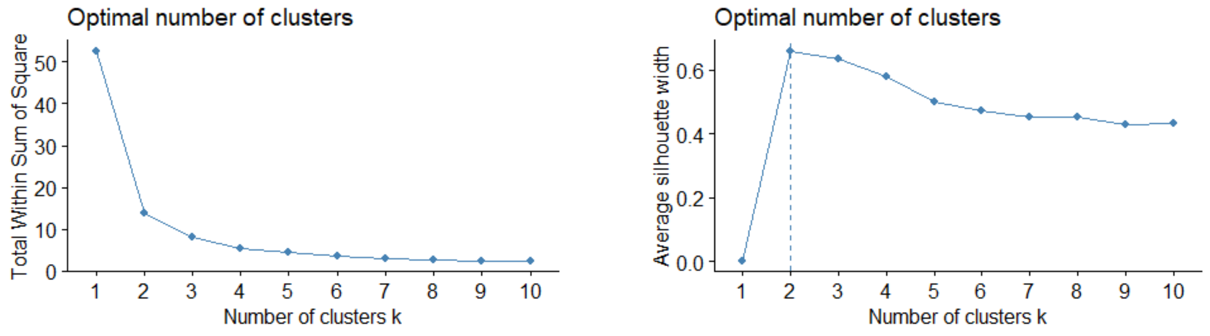

FIGURE 6. ELBOW AND SILHOUETTE GRAPHS FOR THE POKEMON DATASET

Regarding the number of clusters, like in the previous case, the elbow method suggests three clusters, and the silhouette method suggests two as the optimal number of clusters (Figure 6). Three clusters were estimated with the k-means method (Table 3). Similar to the previous case, there are

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

two clusters (2 and 3 in Table 3) in which buffering time values do not seem to be systematically associated with better quality reports (very high p-values, indicating little consistency in the responses of those participants as a function of buffering time). In one of the clusters (cluster 1, containing 101 participants, 56% of the sample), participants appear to be very sensitive to the increase in buffering time: the central value of the cluster is that a 1 standard deviation variation in buffering time produces a 0.78 reduction in the quality measure (MOS), with a p-value close to statistical significance.

Table 3. Clusters obtained from the 'Pokemon' dataset (Amour et al., 2015)

Cluster number	Variable means		N users in cluster
	"coef. buffering time"	"p-value"	
1	-0.78	0.08	101 (61%)
2	-0.14	0.71	42 (25%)
3	0.42	0.40	23 (14%)

The LPA analysis (Figure 7) shows, however, that in this case there are two classes of users that overlap considerably in their sensitivity to buffering time and in the p-values associated with it. This suggests that possibly these two classes of users might be better represented by a single class, so that in the end, the distinction would basically be between two classes of users, those sensitive to buffering time and those who show less sensitivity to it. This indicates that, unlike the previous analysis, users may have a clearer and higher sensitivity in this specific context (as shown also by the higher absolute values in the correlation coefficients).

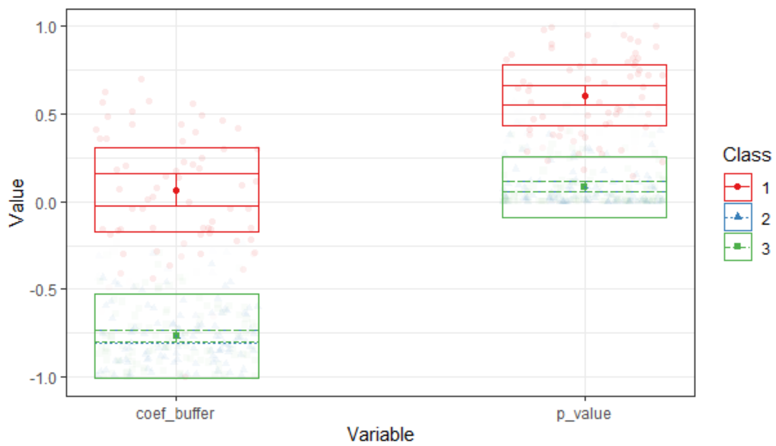

FIGURE 7. RESULTS OF THE LPA ANALYSIS CONDUCTED IN THE POKEMON DATASET

4.1.4.2 Dataset on web browsing QoE

As further validation of the user profiling and additional examination of its reproducibility, a similar analysis was performed with the WebMos-18 dataset (da Hora et al., 2018). This contains 3,010 observations of 181 users performing webpage browsing tasks, with different levels of latency, and their ratings of the quality of their experience were collected on a scale of 1 to 5 (for details, see the original paper (da Hora et al., 2018). As a measure of participant sensitivity, for each participant, the correlation between the system latency ('latency' variable) in the dataset, and the quality ratings given by that user, was analysed. After eliminating users with insufficient data, the final sample contained 161 users. The correlation values and the p-value associated with them, for each user, were used in the cluster analysis.

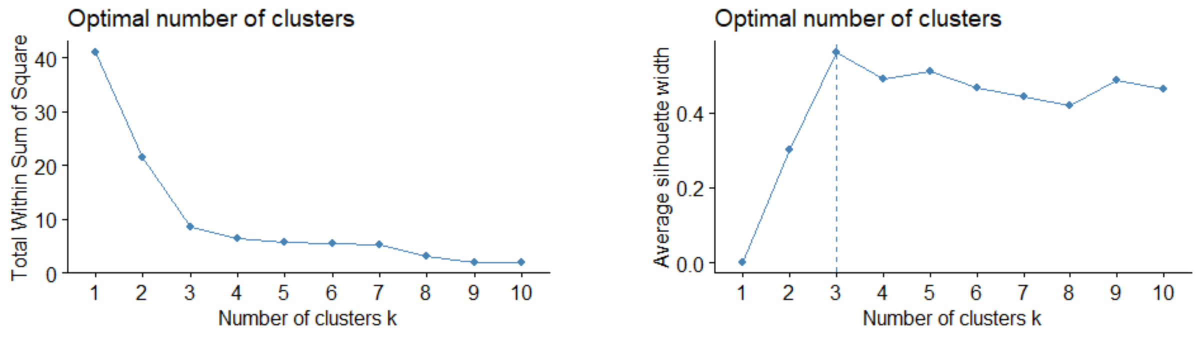

FIGURE 8. ELBOW AND SILHOUETTE GRAPHS FOR THE WEBMOS-18 DATASET

The optimal number of clusters seems to be three in this case (Figure 8). The first of these clusters (cluster 1, Table 4) shows a clear negative correlation between experienced latency and rated quality (although with p-values not very close to statistical significance). This cluster contains 50 participants (31% of the sample). A second cluster contains participants with sensitivity values around zero, and in a third case the sensitivity of the participants seems to go in the opposite direction.

Table 4. Clusters obtained from the "WebMos-18" dataset (da Hora et al., 2018)

Cluster number	Variable means		N users in cluster
	"coefficient"	"p-value"	
1	-0.48	0.13	50 (61%)
2	0.03	0.65	69 (43%)
3	0.49	0.12	42 (26%)

The LPA analysis shows that, in this case (as in the previous analyses, only one group of participants (class 3, in Figure 9) shows a clear and consistent impact (negative coef_lat and p-value close to zero) of latency on their quality of experience. Thus, while there are similarities with the ratings obtained in other contexts, the differences relative to this specific context are also apparent.

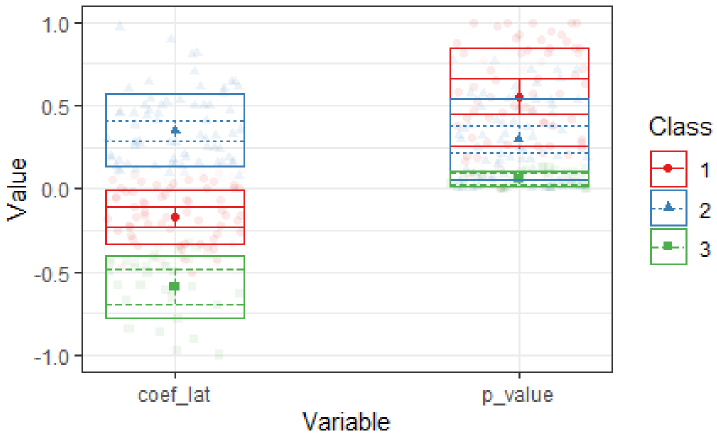

FIGURE 9. RESULTS OF THE LPA ANALYSIS CONDUCTED IN THE WEBMOS-18 DATASET

As a conclusion, the results show consistency in the classification of users into profiles in the three contexts employed, thus demonstrating the relative reproducibility of the results with different data sets. Thus, across the different contexts, the main similarities observed are that the proportion of users showing a clear sensitivity to latency variations is around one third, or 50% at best, and that the impact of latency is always limited and conditioned by other factors (which can be both technology and user related). However, the important differences between contexts (which are consistent with what has been reported in the literature review in previous sections) make explicit the convenience of going beyond the impact analysis on generic measures of quality of experience and addressing the study of its impact on different cognitive dimensions of the user experience, as we address below.

4.2 Analysis the impact of latency on different cognitive and emotional dimensions of user experience

To move towards defining a detailed model of user behaviour in relation to network latency, as we have seen above, it is necessary to address the different cognitive and emotional dimensions that may be involved in user response. Since these are highly context-dependent, it is first necessary to define the experimental use cases in which such impact will be analysed and to implement experimental paradigms to explore them. This section describes **two experimental use cases** for which an experimental methodology has been implemented, the first of which deals with the consumption of streaming videos, while the second deals with the use of online search engines. In addition to defining and implementing the experimental paradigm in both cases, in the first one a

pilot study was conducted with five participants, which allowed us to examine the impact of latency on the cognitive and emotional variables of the participants defined above.

4.2.1 Definition of use cases and experimental paradigms

4.2.1.1 Video streaming

The first of the experimental use cases focuses on the consumption of streaming videos, with the objective of analysing the network latency impacts on indicators of different neuropsychological dimensions of user engagement with the content.

The experimental paradigm designed for this use case is based on studies on QoE in multimedia content, including the implementation of psychophysiological measures. The experimental design includes the manipulation of two independent variables (IVs), each with five levels. The first IV is the latency introduced in the network. For this purpose, six 'qualities' were defined as levels of latency, as well as download and upload speed (Table 5). The different latency levels were determined from levels used in different previous studies, as well as based on the clustering of latency levels observed in real networks. Also, considering that human perception is not linear but exponential, we selected latency values that represent non-linear increases. The rationale for manipulating latency values here is that they can determine aspects of the visual and auditory quality of the video (buffering, bitrate, etc.) in a way that can impact how the video is perceived by users.

Table 5. Levels of 'quality' (based on latency) used in the experiment

'Quality'	Download speed (mbps)	Upload speed (mbps)	Latency (ms)
Q1	87.97	19.32	29.64
Q2	45.22	12.37	54.54
Q3	17.0	6.77	152.31
Q4	19.79	6.53	253.95
Q5	9.3	3.04	658.34
Q6	4.0	1.85	1840.03

To manipulate this IV, we developed a wrapper for Browsertime (https://www.sitespeed.io), a framework designed for automating browser experiments, to render the videos from the YouTube. The wrapper is built with two capabilities: (i) network bandwidth and latency control: Using Throttle (https://www.sitespeed.io), the wrapper regulates network bandwidth and latency, and (ii) automated browser and application actions: It facilitates various automated actions such as maximizing the browser window, enabling video autoplay in YouTube, and setting timers for video playback. Additionally, the wrapper provides an http-based interface using Flask[1], to easily manage and control these capabilities.

The **second independent variable (IV)** considered in this study is **video content**, as this is also key in the perception of quality and could interact with latency. Based on the different contents used in previous studies on multimedia content quality, we selected six YouTube videos representative of five different genres: nature documentary, sports, animation, stand-up comedy, popular science, and music.

Our **experimental design** is a mixed repeated measures design: each participant is asked to view all six contents (one four-minute segment of each). Each of the six contents is presented at one of the six latency levels. The latency level associated with each content rotates among participants, following a Latin square design. Thus, for the whole sample of participants, the different latency levels are presented associated with the different contents (i.e., each content is presented at all latency levels at some point in time). The dependent variables measured in the study include both psychophysiological variables and other variables measured through questionnaires, to capture the different dimensions of the participants' experience. These variables are summarized in Table 6.

Table 6. Dependent variables (DV) included in the study

Dependent variable	Metric
Motivation (motivational approach)	EEG - Frontal alpha assymetry
Visual attention	Occipital alpha
Negative emotional reaction	Parietal beta
Engagement (frontal)	Engagement index - frontal
Engagement (parietal)	Engagement index - parietal
Emotional arousal (tonic)	EDA - Skin Conductance Level
Emotional arousal (phasic)	EDA - Skin Conductance Response
Attentional focus (physiological)	HR & HRV
Negative emotional reaction	EMG - corrugator supercilii

[1] https://flask.palletsprojects.com/en/3.0.x

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Perceived quality (subjective)	Single stimulus continuous procedure (smilar to the one used in Duanmu et al., 2016). A single 100-points scale (from bad to excellent)
Attentional focus (self-reported)	Three items from the Attentional focus subscale in the Narrative Engagement scale (Busselle & Bilandzic, 2009).
Enjoyment (self-reported)	The three items of the Enjoyment Scale (Oliver & Bartsch, 2010)

For the measurement of psychophysiological variables, BitBrain-E32.A1 and BitBrain-BIO.A1 equipment was used. An EEG cap with 32 electrodes was used, as well as electrodes placed on the participant's forehead (EEG) and two fingers of the hand (EDA signal). A photoplethysmography sensor was also used to obtain the participant's HR and HRV during viewing. The presentation of the videos and questionnaires, as well as the sending of marks for the synchronization of the signals was performed using Psychopy[2], while the recording and synchronization of the psychophysiological signals was performed using OpenVibe[3] . The contents were viewed using a ViewSonic VX3276-2K-MHD-2 (32-inch) screen.

4.2.1.2 Use of search engines

Similar to the previous case, an experimental paradigm was also described and implemented to explore the use case of the use of search engines. In this case, the independent variables are, first, the network latency levels (using the same levels and values described in the previous section), while the second IV is the type of search task. We described five search tasks adapted from the literature (Ghosh et al., 2015; Wildemuth et al., 2018). These tasks include searching for information, for five minutes, on topics related to health, wellness, science and technology, e-commerce, and entertainment. These tasks were implemented in Psychopy.

As **dependent variables** in this experimental paradigm, the same psychophysiological variables described in the previous case were included, while validated questionnaires adapted to this use case were used, which are detailed in Table 7.

[2] https://psychopy.org/
[3] http://openvibe.inria.fr/

Table 7. Self-reported dependent variables included in the experimental paradigm on the use of search engines.

Dependent variable	Metric
Overall satisfaction with the results	1-item, 7-point scale scale (Not at all - Totally) (Ararapkis et al., 2021)
Affective state	3 items, 7-point scale - "Bad-Good"; "Tense–Calm"; "Tired–Awake" (Arapakis et al., 2021)
User Satisfaction with the system.	Five items from the Questionnaire for User Interaction Satisfaction (Harper & Norman, 1993): "Terrible–Wonderful"; "Difficult -Easy"; "Frustrating–Satisfy-ing"; "Dull–Stimulating"; "Rigid - Flexible"
User Engagement	Attentional focus and Involvement (3 items each), from the User Engagement Scale by O'Brien & Toms, 2010)
Attentional focus	3 items from the User Engagement Scale (O'Brien & Toms, 2010)
Involvement	3 items from the User Engagement Scale (O'Brien & Toms, 2010)

The implementation of the presentation of the tasks and questionnaires was carried out in Psychopy, while the collection and synchronization of the psychophysiological signals was performed with OpenVibe, in a similar way as described above.

4.2.2 Pilot Study on User Responses to Latency in Video Streaming

A pilot study was conducted using the experimental paradigm defined in section 4.2.1.1. Five participants (male, between 22 and 38 years old) took part in it. The preparation and conduct of the study took about one hour per participant.

4.2.2.1 Data processing and analysis

From the participants' EEG recordings, the different psychophysiological metrics identified in the experimental paradigm (see section 4.2.1.1, Table 6) were calculated for each of the contents viewed by each participant. Metrics based on peripheral psychophysiological measures (tonic and phasic EDA, HR, HRV, and facial EMG) were also employed.

The EEG signal preprocessing was performed using the MNE package in Python[4] and following these steps: (1) a band-pass filter (0.3 - 40 Hz) was applied to eliminate slow drifts; (2) the Fpz channel was taken as EOG, necessary for the use of the algorithm below; (3) each signal was divided into 2 s epochs with 1 s of overlap; (4) the FASTER method (Nolan et al., 2010), based on ICA, to automatically detect, reject and interpolate those epochs with excessively noisy data; and (5) finally, the different EEG-based metrics (FAA, alpha occipital, beta parietal, engagement index - parietal, engagement index - frontal) were calculated.

The EDA signal was processed using the cvxEDA algorithm (Greco et al., 2015), which allowed to break down the signal into its tonic and phasic components. For the processing of the cardiac activity signal and the EMG signal, the Neurokit2[5] package in Python was employed. Finally, each of the signals was divided into epochs with a duration of 1s. Since each video was 4 minutes long, 240 epochs per video were obtained for each participant.

These were analyzed using multilevel mixed models (one per psychophysiological metric). A step-by-step procedure was followed for fitting these models. First, a model was created that included only fixed effects of order of presentation and epochs, to account for possible temporal trends unrelated to latency, as well as a random term for participants (to account for individual differences between participants). Latency fixed effects (expressed as a continuous variable, in ms) were then added to this model. In this way, we explored whether there were any statistically significant linear effects of latency. For the analysis of the questionnaires, a similar approach was followed, although here we considered a single response per video (instead of the multiple epochs per video used in the case of the psychophysiological signals).

4.2.2.2 Results

A summary of the coefficients of the mixed models performed for the EEG metrics is shown in Table 8, while Figure 10 shows the average (normalized) values of each metric, per quality level. As indicated by Table 8, statistically significant results were observed for two metrics: FAA and the frontal region engagement index, while the parietal beta band-based metric yielded marginally significant results. Thus, these results suggest that increased latency levels negatively affect user motivation towards content (indexed by FAA) and may increase the perception of the experience as negative (indexed by the parietal beta metric). This pattern has a clear logic and is consistent with what is expected in this context. However, the increase in the values of engagement indexed in the frontal region is difficult to explain following this logic, although it could be an artefact of the small sample size (i.e. due to high variance), and because it was not possible to fully balance the association of different content with different latency values. Therefore, we argue that content that have been more attractive may have produced higher latency values. On the other hand, no significant changes associated with latency were observed in the measure of visual attention (occipital alpha, or in the engagement index in the parietal region.

[4] https://mne.tools/stable/index.html
[5] https://pypi.org/project/neurokit2/

Table 8. Summary of the coefficients for the models of the EEG metrics

	FAA	Alpha occ.	Beta par.	Engag. Par.	Engag. Front
(Intercept)	0.35***	-8.06***	-8.50***	0.56***	0.57***
order	0.00	-0.02**	0.00***	0.00	0.00
epoch	0.00	0.00***	-0.06***	0.001***	0.00
latency	**-0.06***	0.02	0.03	0.00	**0.01***

* $p < .05$; ** $p < .01$; *** $p < .001$

For readability, the coefficients for latency have been multiplied by 10^3

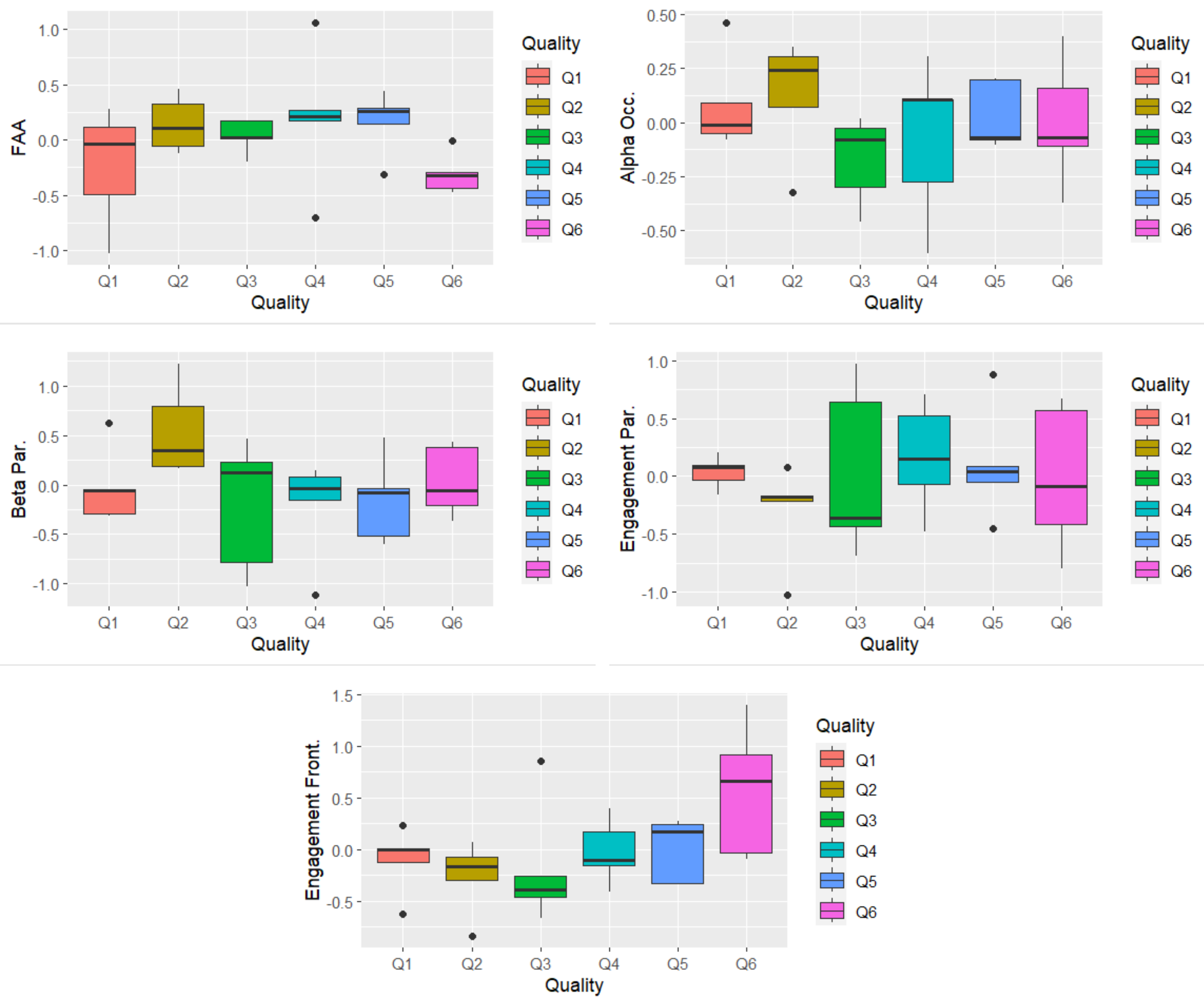

FIGURE 10. BOXPLOTS OF THE NORMALIZED DATA ON EACH EEG METRIC PER QUALITY

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Regarding the peripheral signals (Table 9, Figure 11), an increase in tonic EDA levels associated with latency is also observed. This seems to indicate that higher levels of latency induce higher emotional arousal (which could be related to, for example, a sense of user frustration). This change is only observed at the tonic, rather than phasic, level of EDA, which seems to indicate that this is a sustained aspect during viewing, not limited to single moments. This effect on tonic EDA, understood as an increase in emotional arousal, is also manifested in an acceleration of HR. In turn, this does not seem to impact attentional allocation, indexed by HRV. Increased EMG activity indexes a greater negative valence of experienced emotions associated with higher levels of latency. If this is taken in conjunction with the increased emotional arousal described, there appears to be a clear pattern in which higher latency is related to a higher, negative level of emotion (which would be shareable, for example, with feelings of frustration with the system) associated with the increase in latency.

Finally, as for the results of the questionnaires, we can observe a clear significant impact of latency on the quality reported by users, but not on their level of enjoyment or the attention they say they have paid (Table 10, Figure 12). Thus, users do not consciously report a negative impact of latency on their enjoyment of content. However, this contrasts with the lower motivation (in FAA) and increased negative emotion (in peripheral signals) discussed above, which underscores the usefulness of employing psychophysiological measures to go beyond the information that users can consciously report.

Table 9. Summary of the coefficients for the models of peripheral psychophysiological measures

	EDA - tonic	EDA - phasic	HR	HRV	EMG
(Intercept)	-0.73***	0.11***	72.387***	44.267***	0.541***
epoch	0.00	0.00***	-0.031***	0.011	0.000
order	0.20***	0.01	-0.053	1.223	-0.002
latency	**0.05*** **	0.01	**0.002*** **	0.001	**0.031** **

* *p* < .05; ** *p* < .01; *** *p* < .001

For readability, the coefficients for latency for EDA -tonic, EDSA – phasic, and EMG have been multiplied by 10^3

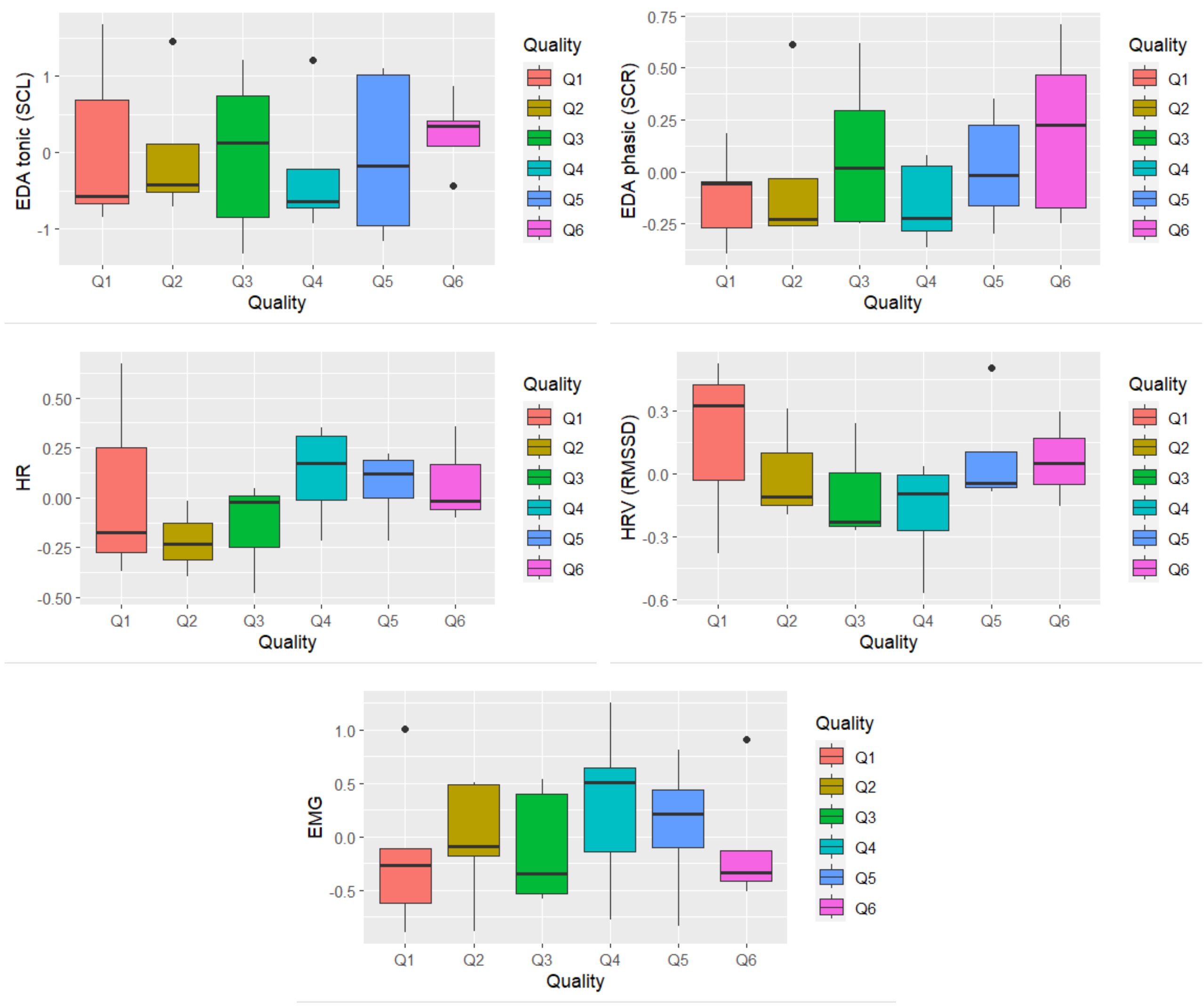

FIGURE 11. BOXPLOTS OF THE NORMALIZED DATA ON EACH PERIPHERAL PSYCHOPHYSILOGICAL MEASURE PER QUALITY

Table 10. Summary of the coefficients for the models of the self-reported measures

	Quality	Enjoyment	Attention
(Intercept)	4.075***	5.41***	4.71***
order	0.014	-0.24	-0.07
latency	**-0.001*** **	0.00	0.00

* *p* < .05; ** *p* < .01; *** *p* < .001

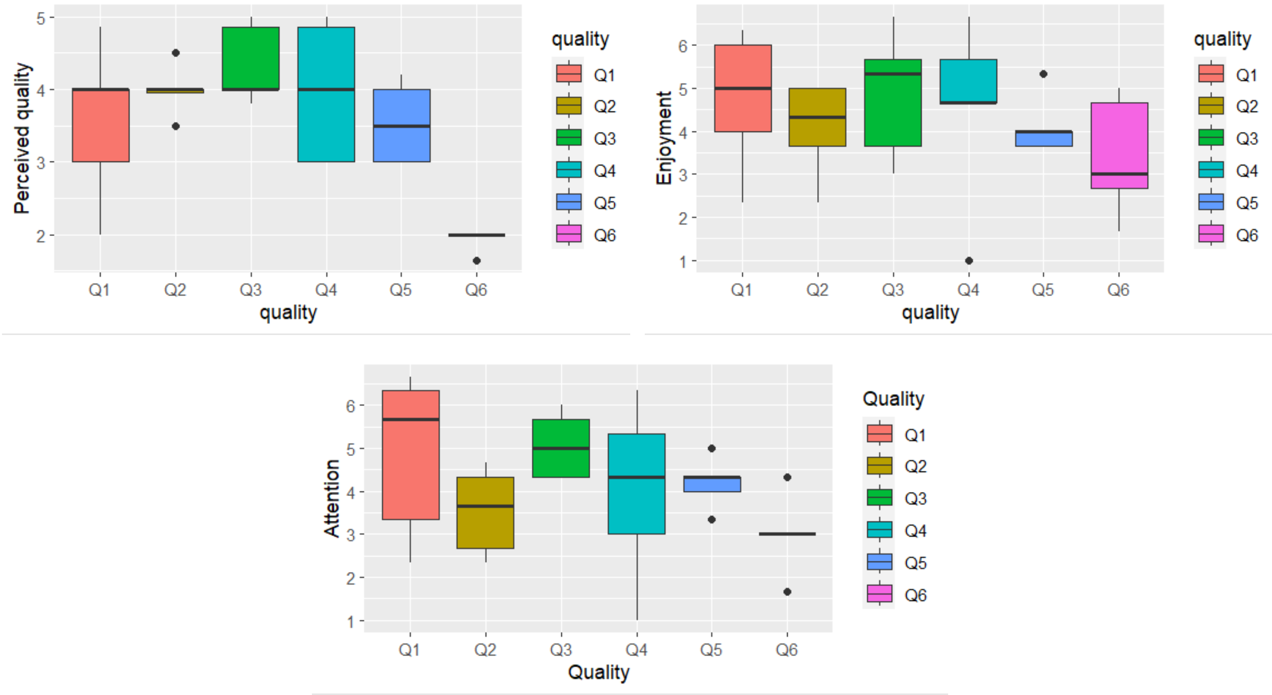

FIGURE 12. BOXPLOTS OF THE NORMALIZED DATA ON SELF-REPORTED MEASURES PER QUALITY

5. Conclusions

To address the objectives of this activity, a multi-method approach was used, combining literature review and conceptual analysis with experimental research based on the use of existing online datasets, the design of experimental paradigms and the collection of data from participants in our laboratory studies.

The first part allowed us to delimit the conceptual and methodological framework of the research, based on the use of concepts and methods from various disciplines, both technical and from the field of social and behavioural sciences. Our methodological approach has revealed important dimensions of the analysis of latency responses. A key aspect here is the difference between the ability to perceive latency and whether latency has an impact on the quality of the experience. Another key insight is the potential impact of latency not only on the conscious subjective assessment of the quality of the experience, but also on other possible aspects of the experience (e.g., the user's motivation, level of attention, emotional response, etc.). Finally, an essential aspect in this regard is that individual differences between users are key in the way they respond to latency. Therefore, the exploration of potential response profiles is a central task.

The review has also identified the channels and sensors suitable for a measurement of the psychological constructs mentioned above. In this regard, the combination of self-reported methods (e.g., questionnaires) with methods based on the use of psychophysiological signals allows a more global understanding of the user's response to latency, going beyond the user's own conscious perception and with a high temporal resolution.

Based on the variables and indicators identified, we have defined an approach to explore the sensitivity of users to latency, capitalizing on the existence of several datasets available online with a large number of users. The algorithm defined for this purpose is based on two steps: a first analysis of the individual sensitivity of each user to latency, and secondly, the use of these data to group users into clusters. This approach has been employed on different datasets, showing that, while the basic three-cluster structure appears to be robust and consistent across different contexts, there is also an important source of context-dependent variability in the clusters. To contribute to a better understanding of this, the next step has focused on going beyond the subjective quality reports present in the datasets used, and employing the psychophysiological responses of the participants to obtain a picture of the different dimensions of the experience that are impacted by latency (including user motivation, level of attention and engagement, etc.). We have defined two experimental paradigms, based on two use cases (streaming video consumption and search engine usage), and conducted a pilot study using the first one, with five participants performing the viewing in our lab. The results have confirmed the impact of latency on lower viewer motivation towards the content, as well as a presence of more negative emotions. These results were evident in psychophysiological measures, but not in some of the questionnaires, which supports the usefulness of using these measures. The next stage of the project will address the use of these measures to refine the profiling of users by conducting more detailed profiles with more dimensions.

Financiado por la Unión Europea NextGenerationEU · Gobierno de España · Ministerio de Asuntos Económicos y Transformación Digital · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales · Plan de Recuperación, Transformación y Resiliencia · UNICO I+D

References

Amour, L., Sami, S., Hoceini, S., & Mellouk, A. (2015, November). Building a large dataset for model-based QoE prediction in the mobile environment. In *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems* (pp. 313-317).

Annett, M., Anderson, F., Bischof, W. F., & Gupta, A. (2014). The pen is mightier: Understanding stylus behaviour while inking on tablets. In *Graphics Interface* 2014 (pp. 193-200). AK Peters/CRC Press.

Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy, 19*(9), 452.

Arapakis, I., Barreda-Angeles, M., & Pereda-Baños, A. (2017). Interest as a proxy of engagement in news reading: Spectral and entropy analyses of EEG activity patterns. *IEEE Transactions on Affective Computing, 10*(1), 100-114.

Arapakis, I., Park, S., & Pielot, M. (2021, March). Impact of response latency on user behaviour in mobile web search. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 279-283).

Attig, C., Rauh, N., Franke, T., & Krems, J. F. (2017). System latency guidelines then and now–is zero latency really considered necessary?. In *Engineering Psychology and Cognitive Ergonomics: Cognition and Design: 14th International Conference, EPCE 2017*, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part II 14 (pp. 3-14). Springer International Publishing.

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience, 20*(3), 327-339.

Barreda-Ángeles, M., Arapakis, I., Bai, X., Cambazoglu, B. B., & Pereda-Baños, A. (2015, August). Unconscious physiological effects of search latency on users and their click behaviour. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 203-212).

Baur, R., Conzelmann, A., Wieser, M. J., & Pauli, P. (2015). Spontaneous emotion regulation: Differential effects on evoked brain potentials and facial muscle activity. *International Journal of Psychophysiology, 96*(1), 38-48.

Boas, D. A., Elwell, C. E., Ferrari, M., & Taga, G. (2014). Twenty years of functional near-infrared spectroscopy: Introduction for the special issue. *Neuroimage, 85*, 1-5.

Bolls, P. D., Weber, R., Lang, A., & Potter, R. F. (2019). Media psychophysiology and neuroscience: Bringing brain science into media processes and effects research. *Media effects: Advances in theory and research,* 195-210.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Boz, E., Finley, B., Oulasvirta, A., Kilkki, K., & Manner, J. (2019). Mobile QoE prediction in the field. *Pervasive and Mobile Computing, 59*, 101039.

Busselle, R., & Bilandzic, H. (2009). Measuring narrative engagement. *Media Psychology, 12*(4), 321-347.

Csikszentmihalyi, M., Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi,* 21-34.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

da Hora, D. N., Asrese, A. S., Christophides, V., Teixeira, R., & Rossi, D. (2018). Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics. In *Passive and Active Measurement: 19th International Conference,* PAM 2018, Berlin, Germany, March 26–27, 2018, Proceedings 19 (pp. 31-43). Springer International Publishing.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. *Handbook of psychophysiology, 2*, 200-223.

De Silva, R. N., Cheng, W., Ooi, W. T., & Zhao, S. (2010, June). Towards understanding user tolerance to network latency and data rate in remote viewing of progressive meshes. In *Proceedings of the 20th International Workshop on Network and Operating Systems Support for digital audio and video* (pp. 123-128).

Debarba, H. G., Montagud, M., Chagué, S., Herrero, J. G. L., Lacosta, I., Langa, S. F., & Charbonnier, C. (2022). Content format and quality of experience in virtual reality. *Multimedia Tools and Applications*, 1-26.

Díaz-Oreiro, I., López, G., Quesada, L., & Guerrero, L. (2019). Standardized questionnaires for user experience evaluation: A systematic literature review. *UCAmI 2019*, 14.

Doherty, R. A., & Sorenson, P. (2015). Keeping users in the flow: Mapping system responsiveness with user experience. *Procedia Manufacturing, 3*, 4384-4391.

Duanmu, Z., Zeng, K., Ma, K., Rehman, A., & Wang, Z. (2016). A quality-of-experience index for streaming video. *IEEE Journal of Selected Topics in Signal Processing, 11*(1), 154-166.

Durnez, W., Zheleva, A., Claypool, M., Maes, M., Bombeke, K., Van Looy, J., & De Marez, L. (2021). Spaz! The Effects of Local Latency on Player Actions in a Desktop-Based Exergame. *IEEE Transactions on Games, 14*(4), 623-631.

Fischer, J. E. (2009). Experience-sampling tools: a critical review. *Mobile living labs 09: Methods and tools for evaluation in the wild*, 35.

Forch, V., Franke, T., Rauh, N., & Krems, J. F. (2017, July). Are 100 milliseconds fast enough? Characterizing latency perception thresholds in mouse-based interaction. In *19th International Conference on Human-Computer Interaction*, Vancouver, Canada (pp. 9-14).

Galloso, I., Palacios, J. F., Feijóo, C., & Santamaría, A. (2016). On the influence of individual characteristics and personality traits on the user experience with multi-sensorial media: an experimental insight. *Multimedia Tools and Applications, 75*, 12365-12408.

Ghosh, S., Rath, M., & Shah, C. (2018, March). Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 22-31).

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2015). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering, 63*(4), 797-804.

Halbhuber, D., Henze, N., & Schwind, V. (2021). Increasing player performance and game experience in high latency systems. *Proceedings of the ACM on Human-Computer Interaction, 5*(CHI PLAY), 1-20.

Halbhuber, D., Schauhuber, P., Schwind, V., & Henze, N. (2023). The Effects of Latency and In-Game Perspective on Player Performance and Game Experience. *Proceedings of the ACM on Human-Computer Interaction, 7*(CHI PLAY), 1308-1329.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Harper, B. D., & Norman, K. L. (1993, February). Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. In *Proceedings of the 1st Annual Mid-Atlantic Human Factors Conference* (Vol. 224, p. 228). sn.

International Telecommunication Union [ITU] (2019). *Recommendation ITU-T P.10/G.100. Vocabulary for performance, quality of service and quality of experience*. Amendment 1.

International Telecommunications Union [ITU] (2021). *Recommendation ITU-T G.1035. Influencing factors on quality of experience for virtual reality services.*

International Telecommunications Union [ITU] (2019). *Recommendation ITU-R BT.500-14 (10/2019). Methodologies for the subjective assessment of the quality of television images.*

Jacques, R. D. (1996). *The nature of engagement and its role in hypermedia evaluation and design* (Doctoral dissertation, South Bank University).

Jota, R., Ng, A., Dietz, P., & Wigdor, D. (2013, April). How fast is fast enough? A study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2291-2300).

Kaaresoja, T., Brewster, S., & Lantz, V. (2014). Towards the temporally perfect virtual button: Touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions. *ACM Transactions on Applied Perception* (TAP), 11(2), 1-25.

Khan, R. A., Naseer, N., & Khan, M. J. (2019). Drowsiness detection during a driving task using fNIRS. In *Neuroergonomics* (pp. 79-85). Academic Press.

Lalmas, M., O'Brien, H., & Yom-Tov, E. (2022). *Measuring user engagement*. Springer Nature.

Laurel, B. (1993). *Computers as Theatre*. ed. Reading: Addison-Wesley Publishing Company.

Le Callet, P., Möller, S. & Perkis, A., (2013). *Qualinet White Paper on Definitions of Quality of Experience* (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003). Version 1.2. Mar-2013

Liu, S., Claypool, M., Kuwahara, A., Sherman, J., & Scovell, J. J. (2021, May). Lower is better? The effects of local latencies on competitive first-person shooter game players. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Long, M., & Gutwin, C. (2018, October). Characterizing and modeling the effects of local latency on game performance and experience. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (pp. 285-297).

Magidson, J., & Vermunt, J. K. (2002). A nontechnical introduction to latent class models. *Statistical Innovations white paper, 1*, 15.

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal access in the Information Society, 14*, 81-95.

Martinek, R., Ladrova, M., Sidikova, M., Jaros, R., Behbehani, K., Kahankova, R., & Kawala-Sterniuk, A. (2021). Advanced bioelectrical signal processing methods: Past, present and future approach—part ii: Brain signals. *Sensors, 21*(19), 6343.

Mauss, I. B., & Robinson, M. D. (2010). Measures of emotion: A reviews. *Cognition and Emotion,* 109-137.

Nardelli, M., Greco, A., Sebastiani, L., & Scilingo, E. P. (2022). ComEDA: A new tool for stress assessment based on electrodermal activity. *Computers in Biology and Medicine, 150,* 106144.

Nardelli, M., Vanello, N., Galperti, G., Greco, A., & Scilingo, E. P. (2020). Assessing the quality of heart rate variability estimated from wrist and finger ppg: A novel approach based on cross-mapping method. *Sensors, 20*(11), 3156.

Ng, A., Annett, M., Dietz, P., Gupta, A., & Bischof, W. F. (2014, April). In the blink of an eye: Investigating latency perception during stylus interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1103-1112).

Ng, A., Lepinski, J., Wigdor, D., Sanders, S., & Dietz, P. (2012, October). Designing for low-latency direct-touch input. In *Proceedings of the 25th annual ACM symposium on User Interface Software and Technology* (pp. 453-464).

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231.

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods, 192*(1), 152-162.

Nuamah, J. K., & Seong, Y. (2018). Support vector machine (SVM) classification of cognitive tasks based on electroencephalography (EEG) engagement index. *Brain-Computer Interfaces, 5*(1), 1-12

O'Brien, H. (2016). Theoretical perspectives on user engagement. *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media*, 1-26.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59*(6), 938-955.

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50-69.

Olbrich, S., Mulert, C., Karch, S., Trenner, M., Leicht, G., Pogarell, O., & Hegerl, U. (2009). EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. *Neuroimage, 45*(2), 319-332.

Oliver, M. B., & Bartsch, A. (2010). Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research, 36*(1), 53-81.

Perrey, S. (2008). Non-invasive NIR spectroscopy of human brain function during exercise. *Methods, 45*(4), 289-299.

Quesenbery, W. (2014). The five dimensions of usability. In *Content and complexity* (pp. 93-114). Routledge.

Raake, A., & Egger, S. (2014). Quality and quality of experience. In Möller, S., Raake, A. (eds) *Quality of Experience. T-Labs Series in Telecommunication Services*. Springer, Cham. https://doi.org/10.1007/978-3-319-02681-7_6

Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing, 44*, 1031-1051.

Reiter, U., Brunnström, K., De Moor, K., Larabi, M. C., Pereira, M., Pinheiro, A., ... & Zgank, A. (2014). Factors influencing quality of experience. In Möller, S., Raake, A. (eds) *Quality of Experience. T-Labs Series in Telecommunication Services*. Springer, Cham. https://doi.org/10.1007/978-3-319-02681-7_6

Rosenberg, J. M., van Lissa, C. J., Beymer, P. N., Anderson, D. J., Schell, M. J. & Schmidt, J. A. (2019). *tidyLPA: Easily carry out Latent Profile Analysis (LPA) using open-source or commercial software* [R package]. https://data-edu.github.io/tidyLPA/

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.

Seshadrinathan, K., Soundararajan, R., Bovik, A. C., & Cormack, L. K. (2010). Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing, 19*(6), 1427-1441.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health, 258.*

Singh, S., Dijkstra-Soudarissanane, S., & Gunkel, S. (2022). Engagement and quality of experience in remote business meetings: A Social VR study. In *Proceedings of the 1st Workshop on Interactive eXtended Reality* (pp. 77-82).

Smith, M. E., & Gevins, A. (2004). Attention and brain activity while watching television: Components of viewer engagement. *Media Psychology, 6*(3), 285-305.

Smith, E. E., Reznik, S. J., Stewart, J. L., & Allen, J. J. (2017). Assessing and conceptualizing frontal EEG asymmetry: An updated primer on recording, processing, analyzing, and interpreting frontal alpha asymmetry. *International Journal of Psychophysiology, 111*, 98-114.

Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., ... & Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1462), 31-38.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, 103445.

Szafir, D., & Mutlu, B. (2013, April). ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1001-1010).

Tao, X., Chen, Z., Xu, M., & Lu, J. (2019). Rebuffering optimization for DASH via pricing and EEG-based QoE modeling. *IEEE Journal on Selected Areas in Communications, 37*(7), 1549-1565.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience, 26*(37), 9494-9502.

Toet, A., Mioch, T., Gunkel, S. N., Niamut, O., & van Erp, J. B. (2022). Towards a multiscale QoE assessment of mediated social communication. *Quality and User Experience, 7*(1), 1-22.

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR), 50*(6), 1-40.

Varela, M., Skorin-Kapov, L., & Ebrahimi, T. (2014). Quality of service versus quality of experience. In Möller, S., Raake, A. (eds) *Quality of Experience. T-Labs Series in Telecommunication Services.* Springer, Cham. https://doi.org/10.1007/978-3-319-02681-7_6

Webster, J., & Ho, H. (1997). Audience engagement in multimedia presentations. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 28*(2), 63-77.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D

Wechsung, I., & De Moor, K. (2014). Quality of experience versus user experience. In Möller, S., Raake, A. (eds) *Quality of Experience. T-Labs Series in Telecommunication Services.* Springer, Cham. https://doi.org/10.1007/978-3-319-02681-7_6

Wechsung, I., Schulz, M., Engelbrecht, K. P., Niemann, J., & Möller, S. (2011). All users are (not) equal-the influence of user characteristics on perceived quality, modality choice and performance. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop* (pp. 175-186). Springer New York.

Wildemuth, B. M., Kelly, D., Boettcher, E., Moore, E., & Dimitrova, G. (2018). Examining the impact of domain and cognitive complexity on query formulation and reformulation. *Information Processing & Management, 54*(3), 433-450.

Wimmer, R., Schmid, A., & Bockes, F. (2019, May). On the latency of USB-connected input devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Wolf, K. (2015). Measuring facial expression of emotion. *Dialogues in Clinical Neuroscience, 17*(4), 457-462.

Zhu, Y., Heynderickx, I., & Redi, J. A. (2015). Understanding the role of social context and user factors in video quality of experience. *Computers in Human Behavior, 49*, 412-426.

Zuo, X., Yang, J., Wang, M., & Cui, Y. (2022, May). Adaptive bitrate with user-level QOE preference for video streaming. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications* (pp. 1279-1288). IEEE.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE TELECOMUNICACIONES E INFRAESTRUCTURAS DIGITALES
Plan de Recuperación, Transformación y Resiliencia
UNICO I+D