The influence of context on the user experience of an interactive television interface: A comparative study between laboratory and ecological settings

Rayane Mabrouki[1,2], Laetitia Gros[1], Julian Moreira[1], Jean-François Petiot[2], & Franck Mars[2]
[1]Orange Innovation, [2]Laboratoire des Sciences du Numérique de Nantes France

Abstract

This study explores the impact of environmental context on user experience (UX) by testing a TV interface prototype in two distinct settings: a user testing laboratory and an open meeting space. These environments differ in objective criteria such as luminosity, seat type, acoustics, and the presence of employees passing by. Sixteen participants took part in both conditions, completing the AttrakDiff2 questionnaire, which assesses pragmatic and hedonic qualities of UX, and a modified User eXperience Context Scale (UXCS-R), designed to evaluate perceived context. The global UXCS-R scores showed no significant differences between the two environments. Consistently, the AttrakDiff scores did not reveal significant variations. The lack of difference in overall context appreciation suggests that UX outcomes were likely to remain consistent across contexts. However, significant correlations were found between multiple dimensions of the two questionnaires (e.g., between the social component of the UXCS-R and the global AttrakDiff score as well as three of its dimensions), indicating an influence of context appreciation on UX, despite similarly appreciated contexts overall. The results are discussed in terms of how emotional responses and context perception influence user judgment and mindset. Future research should use environments where participants' appreciation varies significantly between contexts.

Introduction

Context in user experience

Experience is coloured by the context in which it takes place (Lallemand & Koenig, 2020). Prominent user experience (UX) models underscore the fundamental role of context in shaping the quality of interaction between users and interactive systems (Hassenzahl & Tractinsky, 2006; Mahlke & Thüring, 2007; Roto et al., 2011). The importance of context is recognised, as demonstrated in surveys of both practitioners and researchers of the field (Lallemand, Gronier & Koenig, 2015; Law et al., 2009). However, empirical studies investigating the effects of context on UX remain sparse, and the findings available are sometimes contradictory.

In D. de Waard, L. Onnasch, M. Liebherr, A. Dettmann, F. Siebert, K. Karrer-Gauß, S. Winkler, A. Toffetti, and M. De Angelis (2025). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2025 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Related work

In the Human-Computer Interaction field, the influence of context has been studied in relation to the evaluation of (i) audiovisual content and (ii) interactive systems. Regarding the evaluation of audiovisual content (i), some studies have examined the effect of general contextual differences. For instance, Jumisko-Pyykkö and Hannuksela (2008) and Jumisko-Pyykkö and Utriainen (2011) investigated the influence of four distinct environments—train station, bus, café, and laboratory—on the evaluation of audiovisual content with varying quality levels. They found that content was perceived as less entertaining on a bus compared to other settings. Furthermore, high-quality video streams were rated as more satisfying in a laboratory, while the lowest-quality streams were perceived as less satisfactory in the laboratory compared to other contexts. Other studies, such as those by Zhu et al. (2015), highlighted that the presence of friends positively influenced the appreciation of audiovisual content. Additionally, participants' interest in particular types of content was found to correlate with their evaluations, but only in the absence of co-viewers. Some research has focused on more specific contextual variables. For example, Porcu et al. (2019) demonstrated a negative effect of ambient noise and a positive effect of diffuse lighting (compared to direct lighting) on the audiovisual experience.

In the case of interactive systems (ii), more specific contextual variables have been examined. For example, Barnard et al. (2007) found that light intensity (85 or 260 lux) and the requirement to walk or remain seated influenced performance and mental workload during a task involving a stylus-based user interface. Mandryk et al. (2006) observed that UX was rated more positively when participants played a video game against a friend rather than against the computer. Similarly, Sonderegger and Sauer (2009) reported that stress, performance, and emotional states during smartphone interactions were negatively affected by the presence of passive observers (introduced as the smartphone's designers). Other studies, such as those by Cocchia et al. (2024), showed that population density in public spaces influenced social acceptability but did not alter the UX during an augmented reality gaming session on a smartphone. Sonderegger et al. (2019) explored the influence of global contexts on interactions with an application by manipulating room settings (living room vs. office) and task instructions (organising a social outing with friends vs. preparing for a professional meeting). Their results revealed no significant differences in affective states, mental workload, or perceived usability between these two contexts, findings that align with earlier studies (Sonderegger et al., 2014).

These studies illustrate the influence of contextual elements, whether general or specific, on variables related to the quality of experience (QoE, or perceived content quality) and interaction evaluation. However, research on the effects of general context elements on UX and ergonomics-oriented variables often shows limited context influence.

Ecological validity of user testing

Context is also studied through the lens of ecological validity. While controlled environments are preferred for scientific rigour and control, experimental sciences increasingly adopt naturalistic approaches, described by Rogers (2011) as "in-the-wild". To assess the validity of such methods, some studies compare human-computer interactions between real-world and laboratory settings (e.g., Kjeldskov et al., 2004). Others tried to study this difference by simulating real-world settings by reproducing them in controlled laboratories (e.g., Nielsen et al., 2006) or by using immersive technologies like virtual reality (e.g., Voit et al., 2019).

However, the difference between ecological and laboratory conditions cannot be attributed to specific elements: mostly general context elements are manipulated (ecology of the whole setting), and often the only specific change mentioned is the lack of control (Kjeldskov & Skov, 2014). Moreover, few studies have focused on the influence of ecology on user judgment and UX, but rather more on user and system behaviour (i.e., usability).

Subjective context evaluation

To further explore the relationship between general context changes and UX, Lallemand and Koenig (2020) proposed a questionnaire enabling the subjective evaluation of context by users. This *User Experience Context Scale* facilitates the identification of contextual variables that participants perceive as differentiating between general contexts. This can improve the lack of interpretability identified previously. The impact, or lack thereof, of a location, room, or environment on UX can thus be linked to more specific components or elements of that general context.

Objectives

The present experimental study has three main objectives: (i) to examine the influence of context on UX along two underexplored axes: (a) in the use of an interactive service and (b) in connection with the subjective evaluation of contexts by participants; (ii) to address concerns of ecological validity comparing two interaction environments, a laboratory room and a simulated living room; (iii) to propose methodological recommendations in conducting user testing.

Method

Use case

The employed interactive system was a television service prototype. This system was selected on the basis of several considerations: (i) it offered a variety of interactions that allowed for the design of two distinct scenarios, thereby reducing learning effects between sessions in each context; (ii) the addition of a passive viewing phase made the evaluation of audiovisual quality relevant, another aspect of experience potentially affected by context; (iii) the possibility of spontaneous navigation, without instructions, promoted an assessment detached from purely ergonomic aspects and fostered greater immersion in the environment.

Experimental conditions

Interaction occurred under two conditions, *Lab* and *Bréhat* (see Figure 1). The *Lab* condition took place in a laboratory room designed for audiovisual quality testing, compliant with the International Telecommunication Union standards (ITU-T, 1998). This environment features cool colours (white, blue, grey) and acoustically optimised properties for focused listening. Sound reverberation is minimised by closed curtains, a perforated ceiling, carpeted flooring, and fabric-covered walls. The room is also sound insulated. Adjustable lighting was set at 30 lux. Participants sat on a swivel office chair facing a table with the computer used for interacting with the TV application prototype.

The *Bréhat* condition took place in a meeting area called "Bréhat", opened onto a corridor leading to employee offices. It features modern Scandinavian-style furniture, providing a spacious, comfortable, and naturally lit environment. Unlike *Lab*, *Bréhat* has large windows allowing natural light, with illumination ranging from 150 lux to over 1000 lux, averaging around 400. Participants sat on a sofa facing a circular coffee table holding the computer for interacting with the TV application. The outdoor view and employee movement introduced occasional distractions, such as cars, gardening, or conversations."

Figure 1. Experimental conditions. "Lab" on the left, "Bréhat" on the right.

Measurements

A demographic and usage questionnaire collected data on age, gender, TV usage frequency, viewing locations (domestic and non-domestic), TV service provider, devices used, and typical co-viewers. These data were only collected during the first session, along with consent forms and instructions. Perceived audiovisual quality, video quality, and audio quality were rated using MOS (Mean Opinion Score) items, each on a 9-point scale (1-Bad to 9-Excellent).

The UX of the TV application was assessed using the French version of the AttrakDiff questionnaire (Lallemand et al., 2015), based on Hassenzahl et al. (2003). This tool includes 28 items across four dimensions: Pragmatic Qualities (PQ) for ergonomic aspects, Hedonic Qualities - Identification (HQI) for self-expression and social image, Hedonic Qualities - Stimulation (HQS) for novelty and enjoyment, and Attractiveness (ATT) for overall appreciation. Each item uses a 7-point semantic differential scale.

Perceived contextual qualities were evaluated through the revised UXCS questionnaire across five components: physical, social, task-related, technical, and temporal (see Table 1). Like AttrakDiff, it uses 7-point semantic differential scales. The original UXCS questionnaire (Lallemand & Koenig, 2020) includes objective and subjective items, objective ones being factual, allowing experimenters to understand a context to which they do not have access (e.g., in the case of uncontrolled remote testing).

Table 1. Revised User Context Scale (UXCS-R). Added items in bold. Underlining to indicate what varied between the two poles of the scale.

While using the system, I was / I had	
PHY1	In an unpleasant-pleasant environment
PHY2	Displeased…- Pleased with the temperature
PHY3	Displeased…- Pleased with the lighting
PHY4	**Displeased…- Pleased with the sound environment**
PHY5	**In an uncomfortable…- comfortable environment**
PHY6	**In an ugly…- beautiful environment**
SOC1	Feeling alone…- related to other people
SOC2	Feeling unsupported…- supported
SOC3	**Bored…- Stimulated by the presence of other people**
SOC4	**Stressed…-Relaxed by the presence of other people**
TASK1	Doing several things simultaneously…- Focusing on the task
TASK2	Often…- Never interrupted
TASK3	Focused on the product…- on attaining my goals
TECH1	**Unpleased…- Pleased with the network**
TECH2	**Unpleased…- Pleased with the technological environment**
TEM1	Interacting at an uncomfortable…- a comfortable pace
TEM2	It was not the right…- the right moment to use the system

In this experiment, only subjective items were retained. Internal context-related items were excluded as the present work focuses on context elements that refer to the environment (Mabrouki et al., in press), leaving ten original items, to which we added seven new ones:

- For the physical component, three new items evaluated satisfaction with sound ambiance (PHY4), comfort (PHY5), and aesthetics (PHY6).

- For the social component, two new items assessed stimulation and valence linked to others' presence (SOC3 and SOC4).
- For the technical component, two new items concerned satisfaction with network quality (TECH1) and the technological environment (TECH2).

The skin conductance level (SCL) was measured, as well as the number of skin conductance responses (SCR), to complement explicit methods. SCL allowed to capture differences in arousal (Lang, 1995) not reflected in the questionnaires and interviews. SCR allowed for the identification of differences in the quantity of emotionally and cognitively significant reactions associated with each session, and thus with each experimental condition (Mandryk et al., 2006). Semi-structured and self-confrontation interviews were conducted to complement participants' judgments on the interface and context.

Procedure

The interaction protocol was structured in four phases, allowing for an in-depth exploration of the application. In (i) the *familiarisation* phase, basic navigation instructions were provided to participants to ensure their understanding of the application's functioning. During (ii) the free *unscripted navigation* phase, participants navigated according to their preferences, simulating a realistic interaction with the TV service. In (iii) the *tasks*, or instruction guided interaction phase, participants received specific oral instructions, leading to the use of various TV application features. The experiment concluded with (iv) the *film watching* phase, during which a ten-minute sequence from Blade Runner (American science fiction film, Ridley Scott, 1982) was watched. Among the films available on the application, this one offered two excerpts featuring a succession of equivalent scenes. The first ten minutes (after the credits) and the subsequent ten both contain scenes of dialogue, exposition, interrogation, and flying car movement. The procedure is illustrated in Figure 2.

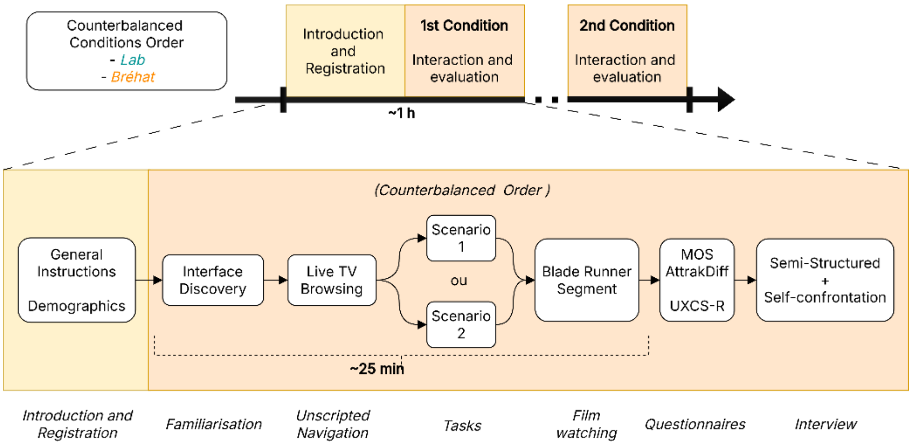

Figure 2. Study procedure.

To avoid a learning effect, the instructions received (and thus the features encountered) by participants during the tasks phase differed and were counterbalanced between the two sessions. The first scenario involved navigating through the replay menu and scheduling the recording of a programme to be broadcast in a few days. The second scenario entailed navigating through the horizontal dropdown menu of live broadcasts, initiating the recording of a live programme, and subsequently deleting a recording. Both scenarios end in the recordings menu of the application to seamlessly transition to the film watching phase.

After removing the headphones, electrodes, and bracelet, participants completed a series of questionnaires on Audiovisual Quality (MOS), User Experience (AttrakDiff), and context appreciation (UXCS-R). Following this, the recorded interview phase began. Upon completion of the interview, participants were thanked, compensated for their participation, and escorted to the exit.

Material

Interaction with the TV application was done via a wired remote control connected to the computer. Sound was delivered through open-back headphones (Beyerdynamic DT 990 PRO 250 ohm), allowing participants to hear both the TV and the experimenter's instructions. Two cameras (Logitech C920 HD Pro) and a microphone (Blue Yeti) recorded the interaction. One camera captured the screen and remote control, while the other recorded facial and bodily expressions. Skin conductance was recorded using a BioNomadix bracelet with two electrodes attached to the middle and index fingers of the participant's non-dominant hand. The bracelet transmitted data to a BIOPAC MP150 acquisition board connected to the experimenter's computer. AcqKnowledge software (5.0) processed the data and synchronised it with recordings from the cameras and audio, facilitating self-confrontation interviews.

Participants

This study involved 16 participants (N = 16, 8 females; 8 males) with a mean age of 43.1 years (SD = 10.3). All of them were recruited from the Orange Innovation Lannion testing panel. Each received gift vouchers worth €60. They read and signed the consent form for the experiment and data collection. In terms of TV service providers, nine reported being subscribed to Orange (closest to the prototype), six to another provider and one to none.

Hypotheses

Three hypotheses were formulated.

H1: Perceived qualities of the environment will influence attributed qualities of the product. This will be reflected in a positive correlation between the UXCS-R and AttrakDiff scores.

H2: Due to better furniture quality, brighter and more open atmosphere, and the relaxed nature of conducting a user test in an open and convivial workspace, participants will perceive the *Bréhat* condition's context more positively. This will be reflected in a higher UXCS-R score, particularly for the physical component.

H3: Therefore, hedonic dimensions of the AttrakDiff questionnaire (HQS and HQI) will be rated more favourably in the *Bréhat* condition compared to the *Lab* condition.

Results

Data analysis

The normality of the data distribution was assessed using the Shapiro-Wilk test. When the data did not follow a normal distribution, the Wilcoxon test was used to compare measures across conditions. For data following a normal distribution, the Student's t-test was applied. For correlations, the Pearson correlation coefficient was used when data followed a normal distribution, and the Spearman correlation coefficient was used otherwise. Results were considered statistically significant when $p < 0.05$.

Correlations between context appreciation and user experience

Several significant correlations were identified between the UXCS-R and AttrakDiff scores (see Table 3). All significant correlations were positive and moderate, ranging between 0.352 and 0.479. The correlation between the overall UXCS-R score (UXCS_R_Total) and both Global Attractiveness (ATT) and Hedonic Qualities - Stimulation (QHS) was significant (respectively, $p = 0.048$ and $p = 0.001$). The correlation between the overall UXCS-R score and the overall AttrakDiff score (ATD_Total, $p = 0.051$).

No significant correlations were found between the physical component of the context and the AttrakDiff dimensions. However, the correlation was significant between the social component of the context and all AttrakDiff dimensions, except for Pragmatic Qualities (PQ). The correlation was significant between the task-related component of the context and Hedonic Qualities - Stimulation (HQS). Similarly, the correlation was significant between the temporal component of the context and Hedonic Qualities - Identification (HQI).

*Table 2. Bivariate correlations between UXCS subscale and AttrakDiff. Statistical significance is indicated as follows: * when p < 0.05, ** when p < 0.01 PHY = physical context component, SOC = social context, TASK = task context, TEM = temporal context, TECH = technical context.*

Perceived context Subscale	Test	Total (AD)	PQ	ATT	HQI	HQS
Total (UXCS)	Pearson cor.	.348	.018	.352*	.314	.466**
PHY	Spearman cor.	.156	-.051	.192	.117	.205
SOC	Pearson cor.	.440*	.064	.467**	.462**	.477**
TASK	Spearman cor.	.123	-.167	.083	.043	.479**
TEM	Spearman cor.	.302	.238	.314	380*	-.016
TECH	Spearman cor.	.133	-.061	.185	.071	.322

Context appreciation

Statistical tests revealed no significant differences in the overall UXCS-R score between conditions (t = 0.466, p = 0.648), nor across its dimensions: PHY (W = 55, p = 0.528), SOC (t = 0.593, p = 0.562), TASK (W = 0.506, p = 0.620), TECH (W = 15.5, p = 0.215), and TEM (t = 24.5, p = 0.757). (see Figure 3).

However, significant differences were observed for three specific items of the UXCS-R: PHY4, PHY6, and SOC3 (respectively: "Displeased…- Pleased with the sound environment"; "In an ugly…- beautiful environment"; "Bored… - Stimulated by the presence of others"). In the Lab condition, participants reported being more satisfied with the sound environment (W = 0, p = 0.002) and less bothered by the presence of others (W = 14, p = 0.026) compared to the Bréhat condition. Conversely, participants found the environment in Bréhat to be more aesthetically pleasing than in Lab (W = 0, p < 0.001).

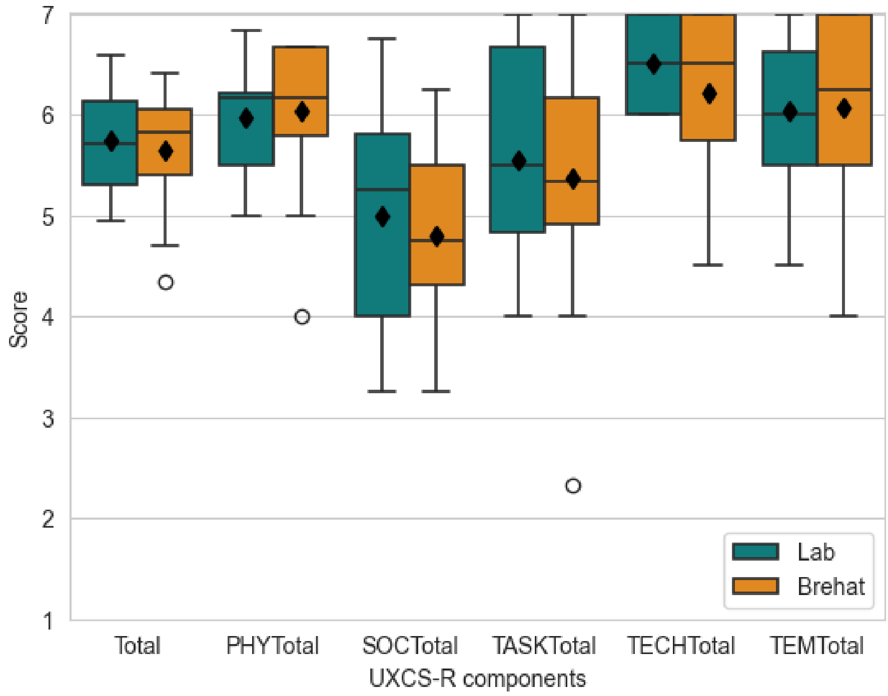

Figure 3. UXCS-R total score and components between conditions. Black diamonds indicate the mean.

User Experience

The overall AttrakDiff score (t = 0.077, p = 0.939) and its dimensions did not differ between the two experimental conditions (see Figure 4): Pragmatic Qualities (t = -0.566, p = 0.580), Attractiveness (t < 0.001, p = 1), Hedonic Qualities - Identification (t = 0.492, p = 0.630), and Hedonic Qualities - Stimulation (t = 0.600, p = 0.557).

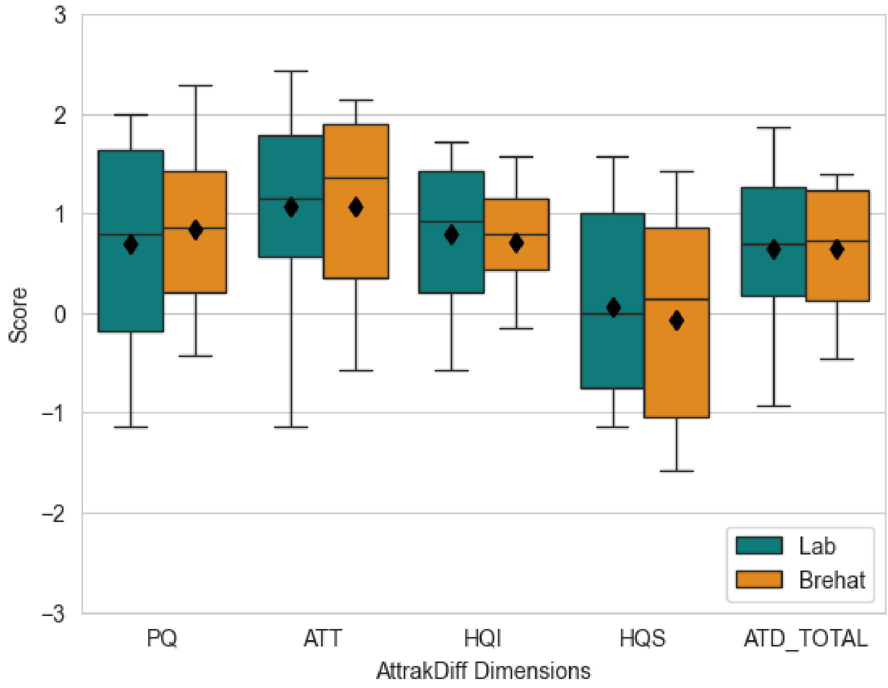

Figure 4: AttrakDiff score by dimension between condition. Black diamonds indicate the mean.

Perceived audiovisual quality

Statistical tests revealed no significant differences for overall Audiovisual Quality (W = 16.5, p = 0.465), Video Quality (W = 36, *p* = 0.501), or Audio Quality (W = 19.5, *p* = 0.403).

UX and usability problems reported

The following result emerged from a post-hoc analysis, not driven by an a priori hypothesis. During the interviews, participants were asked to articulate their opinions on the tested application They were also shown the recordings of their interactions. Criticisms of the application and remote control were coded and counted. The interview analysis revealed a higher frequency of usability problems reported by participants in the Lab condition compared to the Bréhat condition (see Figure 5). This difference was statistically significant (W = 8.5, *p* = 0.006).

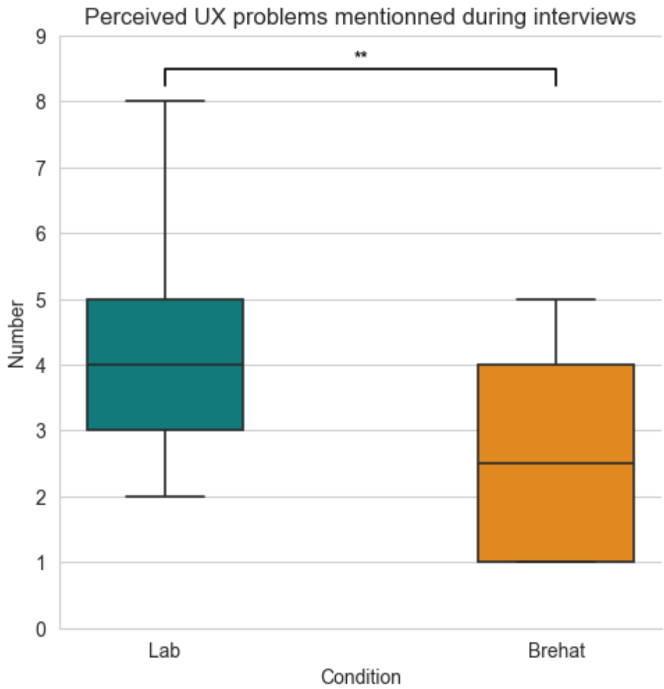

Figure 5. UX problems mentioned during interviews.

Skin conductance level (SCL)

For each session, the average SCL was calculated across the entire interaction protocol as well as for each of its four phases. A trend toward higher SCL, and thus arousal, levels in the Lab condition was identified across the interaction phases: total (W = 19, p = 0.068), familiarisation (W = 22, p = 0.110), unscripted navigation (W = 18, p = 0.057), and tasks (W = 22, p = 0.110). This difference became statistically significant for the film watching phase (W = 16, p = 0.040) (see Figure 6).

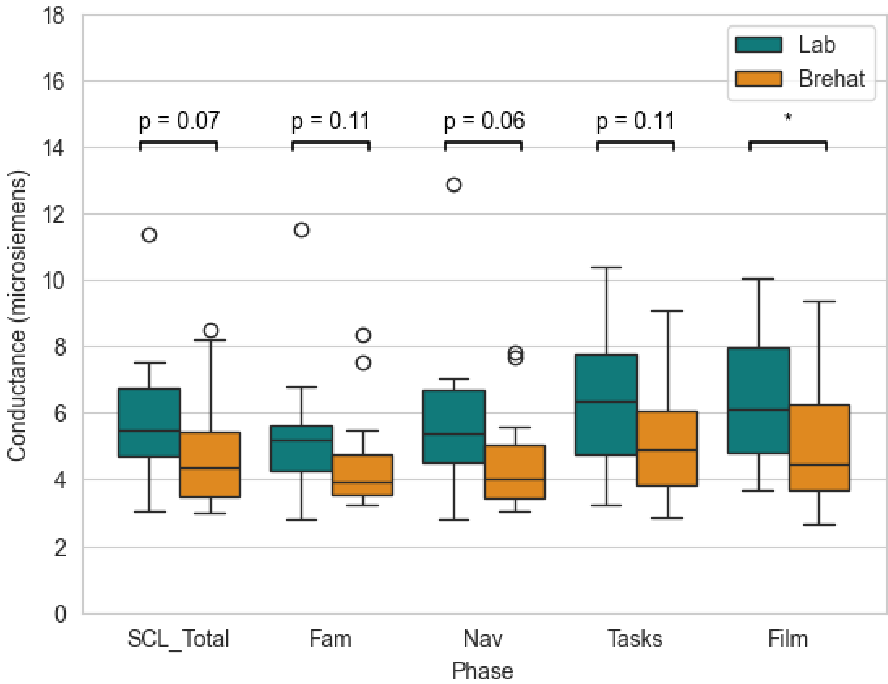

Figure 6. Mean Skin Conductance Level by Phase and Condition.

The data from participants 5, 10, and 11 were excluded from this analysis. For participant 5, the recording quality was insufficient, while for participants 10 and 11, the average values for the periods were anomalously high.

Skin conductance response (SCR)

For each session, the total number of SCR (or peaks) was recorded across the four interaction phases. A slightly higher average number of SCR was observed in the *Bréhat* condition compared to the *Lab* condition, except during the film watching phase. However, this effect was not statistically significant: total (W = 19, p = 0.386), familiarisation (W = 19, p = 0.212), unscripted navigation (W = 38.5, p = 1), tasks (W = 24.5, p = 0.301), film watching (W = -0.450, p = 0.662). Substantial inter-individual variability was observed in these data.

Discussion

Statistical Link Between Context and UX

Consistent with hypothesis H1 and Lallemand and Koenig's (2020) findings, the more positively the context is evaluated, the more the application is considered attractive, and the higher its hedonic qualities are rated. However, this link is not systematic, as not all UXCS-R components correlate with all AttrakDiff dimensions.

This relationship is particularly evident between UX and the social component of the context. This suggest that the more positively the presence of others is perceived, the more favourably the hedonic qualities of the TV application are evaluated. The absence of a link between the social component of context and Pragmatic Qualities suggests that the presence of others did not affect participants' perception of the application's usability. Furthermore, no link is found between any context component and the pragmatic dimension.

It is important to acknowledge the limitations of correlational analyses in establishing causal interpretations. An intermediary variable may affect responses to both the UXCS-R and AttrakDiff questionnaires. Personality traits, mood, or other user characteristics could impact participants' tendency to provide favourable responses to both questionnaires.

Interindividual Differences in Context Appreciation

Contrary to H2, the similar overall scores between conditions suggest a comparable appreciation of both environments. However, analysis reveals antagonistic effects. Items related to noise (PHY4), beauty of the space (PHY6), and boredom caused by others passing by (SOC3), show significant differences, but these differences do not consistently favour the same condition. Additionally, there is a notable interindividual variability.

Several participants mentioned one context resembled their usual TV viewing environment more than the other, but not always the same, and for a variety of reasons. For instance, people passing by in *Bréhat* reminded some participants of their daily interactions with housemates or family (P7, P13, P15), while others associated the sofa and bright lighting with their living room (P6, P8). Conversely, the dim lighting in *Lab* reminded some participants of the ambiance they create at home for watching films (P2, P14). Two participants associated *Lab* with another room, namely their bedroom (P15) and their office (P16). This variability may explain the lack of significant environmental effects on assessments.

Differences in User Experience Based on Testing Environment

Contrary to H3, AttrakDiff scores did not differ between conditions, in concordance with the overall similarity in context appreciation. However, interviews revealed differences in usability issues reported.

The higher number of usability problems reported by participants in the *Lab* condition may be due to the environment's effect on their mindset. Additionally, SCL and self-reported data suggest that participants demonstrated higher levels of concentration and engagement in this condition. These findings may reflect differences in what Hassenzahl refers to as “usage mode.” This concept describes the mindset induced by the situation during interaction, which Hassenzahl et al. (2002) identify as being of two types: goal mode and action mode. Hassenzahl (2003) explains: *In goal mode goal fulfilment is in the fore [...]. Individuals try to be effective and efficient. They describe themselves as “serious” and “planning” [...]. In action mode the action is* *in the fore. [...] Effectiveness and efficiency do not play an important role. Individuals describe themselves as "playful" and "spontaneous".* Participants may have positioned themselves differently along the goal-mode/action-mode continuum depending on the condition. Leaning more towards goal-mode in the *Lab* condition, participants may have focused more on ergonomic aspects of the application. However, the lack of correlation between the Pragmatic Quality dimension of the AttrakDiff and the UXCS-R suggests that the higher number of usability issues reported in Lab may not reflect a greater number of encountered usability problems. Instead, it could simply reflect a tendency to report such issues. This might explain why participants were more critical during interviews in *Lab*, as the formal, serious, and strongly evoking experimentation environment may have encouraged usability feedback compared to the more relaxed Bréhat setting. A higher level of vigilance could also explain the increased arousal. This is supported by the interviews, as eight participants mentioned being more focused in *Lab* or more distracted in *Bréhat*. For *Bréhat*, some participants described the environment as convivial (P2, P13), relaxed (P7), cosy (P5), or reminiscent of a living room (P1, P2, P7, P8, P15, P16).

Conclusion and Outlook

Our findings have several practical and theoretical implications. From a practical standpoint, a more formal environment may be better suited to maximising user feedback and critiques of the evaluated service, as participants reported more usability issues during interviews in the Lab condition. Also, correlations between the social dimension of context and user experience warrants further investigation. This could lead to recommendations regarding the experimenter's role and interaction with participants. From a theoretical standpoint, correlations between UXCS-R and AttrakDiff scores corroborates a link between context and UX. This suggests that contexts must be evaluated more distinctly for a measurable difference in user experience to emerge. Future research could explore a form of halo effect we tried to reveal, namely, associating qualities of the context with the system/product/service being tested.

References

Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., & Moret-Bonillo, V. (2010). A Context-of-Use Taxonomy for Usability Studies. *International Journal of Human–Computer Interaction*, *26*, 941-970.

Barnard, L., Yi, J.S., Jacko, J.A., & Sears, A. (2007). Capturing the effects of context on human performance in mobile computing systems. *Personal and Ubiquitous Computing*, *11*, 81-96.

Cocchia, L., Vergari, M., Kojić, T., Vona, F., Möller, S., Garzotto, F., & Voigt-Antons, J.N. (2024). The impact of social environment and interaction focus on user experience and social acceptability of an augmented reality game. In *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 160-166). IEEE.

Hassenzahl M., Kekez R., Burmester M. (2002). The importance of a software's pragmatic quality depends on usage modes. In L. Holger, A. Çakir; G. Cakir, (Eds) *Proceedings of the 6th international conference on Work With Display Units* (pp 275–276)., ERGONOMIC Institut für Arbeits und Sozialforschung, Berlin.

Hassenzahl, M. (2003). The Thing and I: Understanding the Relationship Between User and Product. In M. Blythe, A. Monk (Eds.), *Funology* (pp. 31-42). Springer.

Hassenzahl, M., Burmester, M., & Koller, F. (2003). AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. *Mensch & Computer 2003: Interaktion in Bewegung*, (pp. 187-196). Vieweg+Teubner Verlag

Hassenzahl, M., & Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, *25*, 91-97.

ITU-T (1998). Recommendation P.911. *Subjective audiovisual quality assessment methods for multimedia applications*.

Jumisko-Pyykkö, S., & Hannuksela, M.M. (2008). Does context matter in quality evaluation of mobile television?. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (pp. 63-72). New York, NY, USA: Association for Computing Machinery.

Jumisko-Pyykkö, S., & Utriainen, T. (2011). A hybrid method for quality evaluation in the context of use for mobile (3D) television. *Multimedia Tools and Applications*, *55*, 185-225.

Kjeldskov, J., & Skov, M. (2014). Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, 43-52. New York, NY, USA: Association for Computing Machinery.

Lallemand, C., Gronier, G., & Koenig, V. (2015). User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in human behavior, 43,* 35-48.

Lallemand, C., & Koenig, V. (2020). Measuring the contextual dimension of user experience: development of the user experience context scale (UXCS). In *Proceedings of the 11th nordic conference on human-computer interaction: shaping experiences, shaping society* (pp. 1-13). New York, NY, USA: Association for Computing Machinery.

Lang, P.J. (1995). The Emotion Probe. *American Psychologist*, *50*, 372-385.

Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 719-728). New York, NY, USA: Association for Computing Machinery.

Mabrouki, R., Gros, L., Moreira, J., Petiot, J.F., & Mars, F. (in press). Approaches of the Context-User-System Triptych – Classifying the Definitions of Context in User Experience Research. In *Proceedings of the 27th International Conference on Human-Computer Interaction (HCII 2025). Lecture Notes in Computer Science vol 15795*. Springer

Mahlke, S., & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 915-918). New York, NY, USA: Association for Computing Machinery.
Mandryk, R.L., Inkpen, K.M., & Calvert, T.W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology*, *25*, 141-158.
Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., & Stenild, S. (2006). It's worth the hassle! the added value of evaluating the usability of mobile systems in the field. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, (pp. 272-280). New York, NY, USA: Association for Computing Machinery.
Porcu, S., Floris, A., & Atzori, L. (2019). Towards the evaluation of the effects of ambient illumination and noise on quality of experience. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE.
Rogers, Y. (2011). Interaction design gone wild: striving for wild theory. *interactions*, *18*, 58-62.
Roto, V., Law, E.C., Vermeeren, A.P., & Hoonhout, J. (2011). User experience white paper: Bringing clarity to the concept of user experience. s.n.
Sonderegger, A., & Sauer, J. (2009). The influence of laboratory set-up in usability tests: effects on user performance, subjective ratings and physiological measures. *Ergonomics*, *52*, 1350-1361.
Sonderegger, A., Uebelbacher, A., Pugliese, M., & Sauer, J. (2014). The influence of aesthetics in usability testing: the case of dual-domain products. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 21-30). New York, NY, USA: Association for Computing Machinery.
Sonderegger, A., Uebelbacher, A., & Sauer, J. (2019). The UX construct–does the usage context influence the outcome of user experience evaluations?. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Proceedings, Part IV 17* (pp. 140-157). Springer International Publishing.
Voit, A., Mayer, S., Schwind, V., & Henze, N. (2019). Online, VR, AR, Lab, and In-Situ: comparison of research methods to evaluate smart artifacts. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-12). New York, NY, USA: Association for Computing Machinery.
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.