A Questionnaire for Collecting Data Relevant to UX Experimental Design

Luka Rukonić[a], Pierre Fastrez[b] and Suzanne Kieffer[c]

Institute for Language and Communication, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Keywords: User Experience, UX Experimental Design, UX Evaluation, Cues, User Scenarios.

Abstract: This paper presents the set of experimental cues involved in the UX experiments that define the characteristics of signals, objects, individuals and prototypes in the lab setting. The contribution of this paper is threefold. First, methodological, as the method employed for creating the questionnaire is reproducible in other domain-applications. Second, practical, as the questionnaire itself can serve as a tool for capturing the experimental cues relevant to the UX evaluation of similar applications. Third, conceptual, as this paper renders a first account of how the questionnaire-collected data can inform other activities ranging from the selection of evaluation methods to the specification of independent variables, UX measures, experimental tasks and apparatus.

1 INTRODUCTION

The development of interactive systems requires User Experience (UX) evaluations with representative end-users, which is an integral part of User-Centered Design (UCD) (ISO, 2010). Typically, UX methods are integrated into the product development lifecycle as a way to follow UCD principles. Within a formative approach, they aim at improving existing design solutions, while within a summative approach, UX evaluations aim at checking whether design solutions meet UX requirements. Also, UX evaluations have a strong impact at the User Interface (UI) level as they always or almost always lead to new or updated User Interfaces (UIs), UX requirements and use cases (Alves et al., 2014). In turn, UX evaluations help to validate design decisions, inform further product development, and achieve UX goals. Because UX is subjective and context-dependent (Law et al., 2009), experimenters need to achieve representative experimental design to allow the generalization of results. It requires capturing the relevant aspects of the real world in order to engage participants in performing the experimental task as they would have for real (Kieffer, 2017). Specifically, experimental designs need to capture both the physical and the digital space of the users, as both spaces are intertwined in the user's experience. The physical space refers to the signals, artifacts and objects typically present in the surrounding of the users, while the digital space refers to the product prototype under investigation. Although necessary for the success of UX evaluations, capturing these data is a complex UX activity, especially in the early stages of the product development lifecycle when the context of use analysis is still underway.

We experienced difficulties capturing these data in an ongoing project involving five other partners and aiming at developing a voice interface for autonomous cars. In this project, we are in charge of the integration of UX into the software development model, an "ad-hoc SCRUM" which combines waterfall and agile methodologies. Previous Human-Computer Interaction (HCI) research (Mayhew, 1999; Maguire, 2001) and UX agile research (Brhel et al., 2015; Garcia et al., 2017) both recommend conducting user requirements analysis or small upfront analysis to extend knowledge about the context of use and user needs before any other development activity. However, conducting such analysis contradicts the agile principle of "changing requirements, even late in development" advocated in the agile manifesto (Beck et al., 2001). Due to the conflicting perspective about user requirements analysis, we had to shrink the analysis process and proceed with design and evaluation processes to align with the agile development team. Although from our perspective it felt like putting the cart before the horse, we had to provide the consortium with experimental designs for UX evaluation without the necessary knowledge about the context of use, that being users, tasks, platforms and environments (Alonso-Ríos et al., 2010).

[a] https://orcid.org/0000-0003-1058-0689
[b] https://orcid.org/0000-0001-7465-4363
[c] https://orcid.org/0000-0002-5519-8814

Rukonić, L., Fastrez, P. and Kieffer, S.
A Questionnaire for Collecting Data Relevant to UX Experimental Design.
DOI: 10.5220/0008164001280136
In *Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications (CHIRA 2019)*, pages 128-136
ISBN: 978-989-758-376-6

This paper reports preliminary findings of using survey research to collect the data relevant for populating the experimental design of the first UX evaluation in our project. We present a questionnaire supporting such data collection, and how we used the collected answers to design the first experiment. In particular, the collected answers allowed us to specify the first instance of our evaluation plan: a selection of the UX evaluation method, specification of independent variables, experimental tasks, scenarios and apparatus.

2 BACKGROUND

2.1 Selecting UX Evaluation Methods

The HCI community has been interested in developing methods supporting UX evaluation (Vermeeren et al., 2010), in building theoretical frameworks to define UX measures (Hassenzahl and Tractinsky, 2006; Law et al., 2009, 2014; Lachner et al., 2016; Zarour and Alharbi, 2017) and developing methodologies for its integration into software development models (Kashfi et al., 2014; Peres et al., 2014). UX evaluation method selection depends on the type of collected data, location of use and phase of development, allowing practitioners to select the most appropriate UX method based on their needs (Vermeeren et al., 2010). However, the current state of UX evaluation. still relies on self-defined questionnaires and post-use evaluation, demonstrating a weak connection between theory and evaluation (Pettersson et al., 2018). Finally, the specification of the experimental tasks involved in UX evaluations remains a choice of the experimenter.

2.2 Conducting Empirical UX Research

The way empirical UX research is conducted is rarely discussed in the HCI literature. The relevant literature on this topic reports that studies either overgeneralize their findings or overemphasize their methodological stance. Regarding the former, while close attention should be paid to experimental design, empirical research in HCI suffers from cause-effect issues and generalization issues (Gray and Salzman, 1998). Regarding the latter, *some studies overemphasize their methodological stance to the extent of damaging research quality. Many uniqueness papers do not report interview questions or protocols, rarely describe data analysis methods, focus mostly on generic UX, and contribute to the dimensionality explosion* (Bargas-avila and Hornbæk, 2011)).

2.3 Designing UX experiments

Designing UX experiments require to control the independent variables to achieve validity of experimental designs. In the context of UX research, such control requires to identify and manipulate the visual, sound and haptic cues that are involved in the experimental design. These cues include the characteristics of the prototypes, signals, objects or individuals that should be available to participants. Therefore, to design UX experiments, researchers need to acquire sufficient knowledge about user goals, needs and limitations, user tasks, platforms used to perform these tasks, and the physical, socio-cultural and organizational environments of users. The relevant literature recommends using UX methods such as surveys, interviews and observations (Maguire, 2001; Law et al., 2009; Vermeeren et al., 2010; Bargas-avila and Hornbæk, 2011; Law et al., 2014).

In turn, how to properly design UX experiments seems to be rarely discussed in the HCI literature and relies on the skills and preferences of UX practitioners and researchers. Previous work (Kieffer, 2017) recommends a seven-step procedure to design UX experiments consisting of task analysis, setting the study goals and metrics, identification of cues, setting the experimental design and assessing its feasibility, assessment of experimental validity and finally conducting the experiment itself and reporting on findings.

The term "cues" refers to traits present in the context of use that users perceive, evaluate and interpret in order to act, react and make decisions (Brunswik, 1956; Araujo et al., 2007; Kieffer, 2017). For example, ice is a visual cue that may indicate slippery roadway risk; beeps are sound cues that may indicate obstacles; steering wheel vibrations are haptic cues that may indicate an unintentional lane change. In the context of UX evaluations, cues refer to any characteristic of the prototypes, signals, objects or individuals available to the participants in the experimental setting. Participants exploit the perceived traits of these cues to understand how to interact with the product prototype, learn its state and how to respond to it.

3 CONTRIBUTION

Because of conflicting perspectives within the consortium about analysis, we were unable to conduct user requirements analysis. Instead, we organized three half-day workshops to specify user scenarios. These workshops were intended for refining the use cases that were written nine months earlier during the writing of our project proposal and for sharing a common big picture of the project goals. However, the project coordinator decided not to proceed with these small upfront analysis activities, as they wanted to move towards development immediately in order to deliver something faster. Paradoxically, the consortium wanted us to design UX experiments, which was impossible because we did not have any information about user needs or goals with the system, what to evaluate or for which purposes.

To work around these issues, we decided to collect the information missing to write relevant experimental scenarios by administrating an online prospective questionnaire about the experimental cues to involve in the UX experiment. The questionnaire is intended for the audience of stakeholders, including product owners, project coordinators and developers. It also serves as a tool for the requirements elicitation from stakeholders as they are the source of information about what scenarios need to be tested when requirements analysis is not done. The contribution of this paper is threefold:

1. Methodological, as the method employed for creating the questionnaire is reproducible in other domain-applications
2. Practical, as the questionnaire itself can serve as a tool for capturing the experimental cues relevant to the UX evaluation of similar applications
3. Conceptual, as this paper renders a first account of how the data collected with the questionnaire can also inform other activities ranging from the selection of evaluation methods to the specification of independent variables and UX measures, experimental tasks and apparatus.

4 METHOD FOR CREATING THE QUESTIONNAIRE

The experimental cues to involved in the UX experiment include the visual, sound and haptic characteristics of the signals, objects, individuals and prototypes available to subjects in the lab setting (Kieffer, 2017).

4.1 Signals and Objects

Signals refer to the stimuli present in the physical environment of the driver such as weather conditions, lighting conditions and level of noise inside and outside the vehicle. Typically, these cues are used by drivers to estimate the comfort and safety while driving, the current state of the vehicle or the situation on the road. Objects such as traffic signals, buildings or vegetation are cues that can be used by drivers to assess the general and local driving conditions: indoor versus outdoor, urban versus rural and the presence of obstacles on the road.

To achieve the representative design and action fidelity (Stoffregen et al., 2003), the experimental design needs to capture these cues and reproduce them in the lab. Examples of real signals and real objects to be reproduced in the lab setting include noise of the engine, honks, sirens, rain drops on the windshield, daylight, traffic signals or obstacles on the road.

4.2 Individuals

This cue consists of a set of traits that describe the social interaction between a driver and other individuals or groups during the experiment such as talking to other passengers in the car, accessing social networks, making phone calls and collaboration with the traffic officers or other types of traffic regulation personnel. The social environment constitutes the presence of other people, collaboration and interaction among them. In other words, *the people with whom the user interacts and who affect the user's interaction with the system* (Alonso-Ríos et al., 2010). In reality, drivers often communicate with other people and interact with various devices which all contribute to the driver's distraction from driving and the increase of their cognitive load. Social, physical and technical environments are the factors that directly influence the usability, design and the use of the system (Maguire, 2001; Alonso-Ríos et al., 2010). Depending on the chosen evaluation method, the experimenter can take the role of an observer, participant or a wizard (e.g. in WOz), performing tasks such as recording users' answers and behavior, taking notes, etc. The experimenter becomes a cue in case there is an interaction happening between the experimenter and the user.

4.3 Prototype

A prototype is the representation of a computer system, characterized by means of five dimensions: visual refinement, interactivity, data model and breadth and width of functionality (Table 1) (McCurdy et al., 2006). The level of fidelity varies across each dimension, supporting mixed-fidelity prototyping. Mixed-fidelity prototyping allows for tailored prototypes to meet specific goals of UX evaluations. For example, the evaluation of the interaction with a prototype requires a rich data model, but the level of visual refinement can be kept low (McCurdy et al., 2006). The level of fidelity of prototypes (low versus high) influences the outcomes of UX evaluations: differences in

Table 1: Independent variables related to the prototype.

Dimension	Values
level of visual refinement	low, medium, high
breadth of functionality	completion (%)
depth of functionality	completion (%)
level of interactivity	low, medium, high
richness of data model	low, medium, high

the nature of usability issues detected (Walker et al., 2002) and differences in the feedback received from participants (Sefelin et al., 2003). The experimenters manually assess the fidelity of the prototype in each of the five dimensions once the prototype is effectively available.

4.4 Resulting Questionnaire

The questionnaire (Table 2) includes 12 items (Q1 to Q12), each item reflecting one independent variable relevant to inform the experimental design from signals, objects and individuals class of cues. Each question starts with the words ”Please specify...”. Items Q1-2 and Q6-9 are closed and multiple-choice questions, whereas items Q3-5 and Q10-12 are open questions. Items Q1-9 are mandatory. We have combined the existing use cases and the description of classes of cues to write the questions.

5 DATA COLLECTION

5.1 Questionnaire Administration

We administered the questionnaire to four industrial partners out of five, which totals 19 potential respondents. We excluded one academic partner from the questionnaire administration, as their mission in the project does not involve any design or development activities. We collected the data from interdisciplinary teams having different needs to develop the final product as they work on the development of various components of the system.

In turn, we collected four survey answers, each representing one partner. It is worth mentioning that these answers represent combined multi-subject responses, as more than one person worked on delivering the answers to the questions. Therefore, more than four persons participated in the survey, but they delivered a unified response reflecting their company's opinion on what kind of scenarios should be tested according to the type of needs they have to successfully develop the final product.

5.2 Results

The results combine quantitative and qualitative findings (Table 3). All respondents indicated that various weather conditions need to be included in the experimental design. None selected the snowy and icy conditions. Among respondents, daytime was the most commonly selected part of the day, followed by night and then dusk or dawn. All respondents answered that the experimental design should involve outdoor (e.g. parking lot or road) and urban setting (e.g. in a city). Three respondents selected medium traffic density (i.e., some cars, fluid traffic), one selected low traffic density (a few cars). Two respondents out of four selected interactions through interactive systems. Finally, all respondents answered that the experimental design should involve social interactions with passengers and no social interactions with individuals on the road (e.g. traffic officers).

6 OUTCOMES

The collected data enabled us to select the UX evaluation method, write testable user scenarios and extract the task sequence models as part of UX evaluation. Furthermore, the prototype fidelity and the instruments used are to be defined by the researchers based on the selected UX evaluation method. For example, it would depend on whether the prototype will require developers to code some components or not, the location of the evaluation or whether the users will be involved (user or expert-based evaluation).

6.1 Selection of the UX Evaluation Method

Given that we will be designing an in-car Voice User Interface (VUI), we decided to use a Wizard of Oz (WOz) technique. Using the classes of cues described above, we designed a WOz experiment to simulate the future system and evaluate the UX while using it. Generally, WOz is used in the early design phase, but it fits well into an iterative design process. In a WOz experiment, ”wizards” simulate only a part, or the whole system while users are interacting with it without being aware that the system is not real (Dow et al., 2005). Nevertheless, WOz experiments can be relatively complex and challenging to implement. The wizards analyze the user's input, determine and generate the output and simulate the behavior of the system. This way the system can be evaluated before it is developed and thus help derive requirements that can then be safely implemented (Bernsen et al., 1993).

Table 2: Questionnaire (soc. interact. = social interaction).

Item	Question: Please specify...	Answer options
Q1	weather conditions	sunny, clear, dry, cloudy, foggy, rainy, snow, icy, windy
Q2	lighting conditions	day, night, dusk, dawn
Q3	time of the day or year	
Q4	noises in the passenger compartment	
Q5	exterior noises	
Q6	general driving conditions	indoor or outdoor
Q7	local driving conditions	urban or rural
Q8	potential obstacles on the road	pedestrians, bicycles, roadworks
Q9	traffic density	none, low, medium, high
Q10	soc. interact. through interactive systems	
Q11	soc. interact. with passengers	
Q12	soc. interact. with individuals on the road	

Table 3: Results.

Item	Collected data
Q1	good (sunny: 2; clear: 3; dry: 2), bad (cloudy: 3; foggy: 1; rainy: 2; snow: 0; icy: 0; windy: 2)
Q2	day: 4; night: 2; dusk: 1; dawn: 1
Q3	no specific time: 4
Q4	ambient noise: 3; engine noise: 1; passengers talking: 2; music: 3
Q5	traffic noise: 1; sirens: 1; engine noise: 1; tyres noise: 1
Q6	indoor: 0; outdoor: 4
Q7	urban: 3; rural: 1
Q8	a pedestrian crossing the road: 1; no obstacles: 1
Q9	none: 0; low: 1; medium: 3; high: 0
Q10	yes: 2; no: 2
Q11	yes: 4; no: 0
Q12	yes: 0; no: 4

The WOz technique has been extensively used in automotive research because it allows for a certain level of improvisation, flexibility, and identification of possibilities of the future system. It has been used to evaluate user expectations, speech-based in-car systems and use of gestures as input modalities in cars (Mok et al., 2015; Lathrop et al., 2004). Furthermore, WOz also served as design research and design prototyping tool allowing researchers to perform remote observation and interaction prototyping of driving interfaces in a car in real time. This way, they could perform contextual interviews with the drivers as well as discover implications for the design of car interfaces (Martelaro and Ju, 2017).

6.2 Generating Scenarios

From the collected questionnaire data, we constructed several testable user scenarios that depict the interaction between the driver and the VUI of an autonomous car. The selection of variables to be included in the experiment is based on their frequency of occurrence in the set of answers. The procedure is as follows:

1. Select the most frequently chosen variables
2. Check for the consistency among variables so that there are no conflicts between them (e.g. sunny and rainy at the same time; day and night; indoor and outdoor)
3. In case of inconsistencies, decide which one will be selected
4. Classify variables into classes of cues
5. Assign them selected values
6. Write a scenario containing the selected variables
7. Create a task sequence model(s)
8. Decide on the prototype characteristics (Table 1)

Following this procedure, researchers can generate several scenarios by either selecting the independent variables based on their frequency of occurrence or by looking at the available data and combining them to come up with a scenario that satisfies the UX evaluation needs. For example, if the goal is to evaluate the system in rainy and dark conditions, then the

Table 4: Independent variables for the User Scenario 1.

Class of cues	Independent variable	Value
Signals	weather	sunny
	lighting	day
	noise	radio on and ambient noise
Objects	general driving conditions	outdoor
	local driving conditions	urban
Individuals	social interaction	yes
	experimenter(s)	no
Prototype	level of visual refinement	medium
	breadth and depth of functionality	low
	level of interactivity	high
	richness of data model	medium

weather variable will be set to "rainy" and the lighting variable to "night".

Previous research used scenarios to represent real-world situations, describe the study design and conditions drivers face on the road. Manawadu et al. (2015) analyzed the driving experience of autonomous and conventional driving using the driving simulator in urban, rural, expressway and parking areas. Scenarios enabled the identification of whether drivers prefer autonomous or conventional driving, depending on the situation and events occurring while driving in different areas. Rödel et al. (2014) used user scenarios to describe the driving situations that correspond to each of the five levels of vehicle autonomy and investigated the relationship between the degree of autonomy and user acceptance and UX factors. Frison et al. (2019) evaluated the influence of road type and traffic volume in different scenarios on the fulfillment of psychological needs during fully automated driving using a simulator and showed that automated driving increases a lack of trust, lowers stimulation and makes drivers want to intervene in driving tasks .

In this paper, we present the "first available spot scenario" and its corresponding task sequence model. It depicts the situation where the driver wants to park the car at any available parking spot on the current route while approaching driver's final destination. Additionally, we defined two more scenarios describing a predefined outcome where the driver can set preferences related to the type of parking place (accessibility, space, distance, etc.), and the collaborative scenario where the driver and the car work together to find a parking place and exchange information meanwhile. We did not include them in this paper, but we derived them using the previously described procedure. All scenarios are related to a semi-automated parking use case.

6.2.1 Scenario 1 - First Available Spot

Alicia is being driven in the autonomous mode. It is a clear and sunny day and she feels like taking a walk. The radio in the car is on. She tells the car to park few blocks away from her apartment. The traffic density is low. The car confirms that the parking search process has started. Alicia wants to meet Bob for coffee in an hour and instructs the car to text him. Eventually, the car has found a parking place but she refuses it. Next, Alicia spots an empty parking place she likes and tells the car to park there. Car confirms that the message has been sent and notifies Alicia to wait for the road to be cleared of passing cars before starting a parking maneuver. Then, Alicia takes over the control of car pedals and with the car's help the parking process finishes. Finally, she leaves the car.

Independent Variables. The independent variables related to the experimental cues of the user scenario 1 are presented in Table 4. These are the values that have to be set to each of the cues involved in the UX experiment and simulated in the lab.

Task Sequence Model. The user scenario was translated into a task sequence model, to enable the identification of user steps, possible errors and breakdowns, as well as triggers and intents that initiate user tasks. This way, we will be able to identify when will the wizard interact with the user (wizard's "hooks"), and when will the response be sent to the user. Moreover, we will be able to compare if the users follow or deviate from the task sequence. The user tasks supported in the experiment are presented in Table 5. The task sequence model for Scenario 1 is shown in Figure 1.

The WOz will help us collect speech data and analyze users' utterances to identify intents and entities used in their dialog. Previous research shows that

Title: Search for a first available parking place to park as soon as possible

Intent: A driver wants to park the car regardless of the location or any other conditions

Trigger: The car has reached the destination

Request the car to park

↓

Tell the car to detect a first free spot close to driver's house

Intent: Park as soon as a suitable parking place is detected

↓

Car starts searching for a parking place

↓

Soon after, the car finds a place and shows it to the driver

↓

⚡ **BD**: The driver refuses because it's too far from the house

↓

Car continues to search for a suitable parking place

↓

The drivers wants to send a message to a friend to set up a meeting

↓

Car looks into contacts and asks for confirmation

↓

The drivers confirms the contact

Intent: the found contact is correct and send a message

↓

Eventually, the car finds another parking place

↓

The drivers confirms the parking place

Intent: I want to start the parking procedure and I like this place

↓

The car starts the parking procedure

↓

⚡ **BD:** There are cars passing by and the car needs to wait for the street to be clear of cars

↓

The car requests the driver to take over the control of the car pedals and gearbox

↓

The drivers confirms that he can take over the control

↓

Control the car pedals and gearbox to park the car

↓

The driver confirms that the parking is finished

Intent: Validate the parking process

↓

Leave the car

Figure 1: Task sequence model for the User Scenario 1.

user utterances vary depending on the driving context and currently performed tasks. For example, utterances spoken while stopped at a traffic light might be more complex than the ones spoken while driving at a curvy road (Lathrop et al., 2004). These intents and entities will feed directly into the design and development of the conversational agent to build dialog flows. Also, WOz will help us evaluate the future system in the early development phase, identify additional user tasks, assess the UA of a voice-controlled autonomous car and evaluate the UX with it.

Table 5: List of user tasks supported in the experiment.

ID	Task name
T1	Request the car to park
T2	Confirm or deny the parking place
T3	Control the car manually (pedals, gearbox)
T4	Show the car a particular parking place
T5	Specify the parking conditions
T6	Make phone calls/send texts
T7	Control the radio/entertainment system

6.3 UX Measures and Instruments

For the UX measurements to be valid and reliable, they need to be collected from a sample of real users, carrying out real tasks in a realistic context of use. However, measuring UX directly and holistically is not possible (Bevan, 2008; Law et al., 2014). Measuring UX seems to be a difficult task that depends on the type of tasks, their timing, methods used, type of information that is collected, etc Law et al. (2014). Clearly, UX measures should provide arguments and implications for the redesign of the evaluated system.

Instruments are the tools used for data collection during UX evaluation. We aim to capture momentary experiences to understand what triggers specific emotions or physiological reactions. The FaceReader enables the real-time measurement of users' emotions which are happiness, anger, disgust, sadness, confusion, fear and neutral state to study UX of single episodes (Vermeeren et al., 2010; Zaman and Shrimpton-Smith, 2006). Thus, the FaceReader will enable us to link user's emotions to actions, utterances or tasks being performed at a specific moment. Finally, we will use the standard questionnaires to collect self-reported data such as task-load and hedonic and pragmatic qualities, the NASA-TLX and AttrakDiff respectively. These data will help us derive requirements that can be safely implemented and that comply with the user's needs.

7 LESSONS LEARNED

Using the data collected with the questionnaire, we produced multiple user scenarios and their task sequence models, but we presented only one. The questionnaire served as a requirements elicitation tool to design the experiment and identify experimental tasks. We realize that the procedure to create scenarios is not deterministic and can yield different scenarios depending on the preferences of the experimenter. However, researchers can generate a set of scenarios following our procedure and choose those that are most valuable for development. Therefore, the experimenter can, depending on the goals of the evaluation, make deliberate choices to study how each cue affects the UX. The UX evaluations need to resemble the real users, real tasks and real environment to increase the ecological validity and obtain representative findings. We deliberately made the questionnaire specific to the project in the automotive sector for the purpose of this case study. Nevertheless, the methodology to create the questionnaire can be replicated to other domain-applications by adapting the questions to a specific domain because the cues are generalizable.

8 CONCLUSION

This paper presents the set of experimental cues involved in the UX experiments that define the characteristics of signals, objects, individuals and prototypes in the lab setting. We used these cues as a methodological basis to construct the questionnaire where each class of cues represents a set of variables. We used this questionnaire to collect the data to form the UX experimental design, create the experimental tasks, select the UX evaluation method and UX measures. The questionnaire served as a tool to work around the missing requirements analysis and helped us to keep up with the development team. The cues represent the independent variables of the experiment which the researchers can use to generate testing scenarios and study how their manipulation affects the results of the UX evaluation. We plan to reuse the questionnaire whenever we need to define the experimental tasks for the UX evaluation to enhance the communication between UX and development teams and understand what types of scenarios need to be tested to properly inform the system's design. Also, the questionnaire allowed us to quickly collect the data necessary to create the UX evaluation plan that we could present to the agile team and align with the development process.

ACKNOWLEDGEMENTS

The authors acknowledge the support by the project VIADUCT under the reference 7982 funded by Service Public de Wallonie (SPW), Belgium.

REFERENCES

Alonso-Ríos, D., Vázquez-García, A., Mosqueira-Rey, E., and Moret-Bonillo, V. (2010). A context-of-use taxonomy for usability studies. *International Journal of Human-Computer Interaction*, 26(10):941–970.

Alves, R., Valente, P., and Nunes, N. J. (2014). The state of user experience evaluation practice. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, NordiCHI '14, pages 93–102, New York, NY, USA. ACM.

Araujo, D., Davids, K., and Passos, P. (2007). Ecological validity, representative design, and correspondence between experimental task constraints and behavioral setting: Comment on Rogers, Kadar, and Costall (2005). *Ecological Psychology*, 19(1):69–78.

Bargas-avila, J. A. and Hornbæk, K. (2011). Old Wine in New Bottles or Novel Challenges? A Critical Analysis of Empirical Studies of User Experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2689–2698.

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., et al. (2001). Manifesto for agile software development. http://www.agilemanifesto.org/. 2019-04-26.

Bernsen, N. O., Dybkjær, H., and Dybkjær, L. (1993). Wizard of Oz prototyping: How and when. In *CCI Working Papers in Cognitive Science and HCI*.

Bevan, N. (2008). Classifying and selecting UX and usability measures. In *International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*, volume 11, pages 13–18.

Brhel, M., Meth, H., Maedche, A., and Werder, K. (2015). Exploring Principles of User-Centered Agile Software Development: A Literature Review. *Information and Software Technology*, 61:163–181.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press.

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J. D., and Gandy, M. (2005). Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing*, 4(4):18–26.

Frison, A.-K., Wintersberger, P., Liu, T., and Riener, A. (2019). Why do you like to drive automated?: A context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, pages 528–537, New York, NY, USA. ACM.

Garcia, A., Silva da Silva, T., and Selbach Silveira, M. (2017). Artifacts for Agile User-Centered Design:

A Systematic Mapping. *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 5859–5868.

Gray, W. D. and Salzman, M. C. (1998). Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-computer interaction*, 13(3):203–261.

Hassenzahl, M. and Tractinsky, N. (2006). User experience-a research agenda. *Behaviour & information technology*, 25(2):91–97.

ISO (2010). Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. Standard, International Organization for Standardization, Geneva, Switzerland.

Kashfi, P., Feldt, R., Nilsson, A., and Svensson, R. B. (2014). Models for integrating UX into software engineering practice: an industrial validation. *Software Engineering*.

Kieffer, S. (2017). ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations. *AIS Transactions on Human-Computer Interaction*, 9(2):149–172.

Lachner, F., Naegelein, P., Kowalski, R., Spann, M., and Butz, A. (2016). Quantified UX: Towards a Common Organizational Understanding of User Experience. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16*, pages 1–10.

Lathrop, B., Cheng, H., Weng, F., Mishra, R., Chen, J., Bratt, H., Cavedon, L., Bergmann, C., Hand-Bender, T., Pon-Barry, H., et al. (2004). A wizard of oz framework for collecting spoken human-computer dialogs: An experiment procedure for the design and testing of natural language in-vehicle technology systems. In *Proc. ITS*.

Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., and Kort, J. (2009). Understanding, Scoping and Defining User Experience: A Survey Approach. *CHI 2009*, 23(1):23–32.

Law, E. L. C., Van Schaik, P., and Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human Computer Studies*, 72(6):526–541.

Maguire, M. (2001). Context of use within usability activities. *International Journal of Human-Computer Studies*, 55(4):453–483.

Manawadu, U., Ishikawa, M., Kamezaki, M., and Sugano, S. (2015). Analysis of individual driving experience in autonomous and human-driven vehicles using a driving simulator. In *2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 299–304. IEEE.

Martelaro, N. and Ju, W. (2017). WoZ Way: Enabling Real-Time Remote Interaction Prototyping & Observation in On-Road Vehicles. *Accepted: Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pages 169–182.

Mayhew, D. J. (1999). *The usability engineering lifecycle: a practitioner's handbook for user interface design*. Morgan Kaufmann.

McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, B., and Vera, A. (2006). Breaking the Fidelity Barrier - An Examination of our Current Characterization of Prototypes and an Example of a Mixed-Fidelity Success. *Proceedings of the International Conference on Human Factors in Computing Systems (CHI'06)*, pages 1233–1242.

Mok, B. K.-J., Sirkin, D., Sibi, S., Miller, D. B., and Ju, W. (2015). Understanding Driver-Automated Vehicle Interactions Through Wizard of Oz Design Improvisation. *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: driving assessment 2015*, pages 380–386.

Peres, A. L., Silva, T. S. D., Silva, F. S., Soares, F. F., Carvalho, C. R. M. D., and Meira, S. R. D. L. (2014). Agileux model: Towards a reference model on integrating ux in developing software using agile methodologies. In *Proceedings of the 2014 Agile Conference*, AGILE '14, pages 61–63, Washington, DC, USA. IEEE Computer Society.

Pettersson, I., Lachner, F., Frison, A.-K., Riener, A., and Butz, A. (2018). A Bermuda Triangle?: A Review of Method Application and Triangulation in User Experience Evaluation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA. ACM.

Rödel, C., Stadler, S., Meschtscherjakov, A., and Tscheligi, M. (2014). Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '14, pages 1–8, New York, NY, USA. ACM.

Sefelin, R., Tscheligi, M., and Giller, V. (2003). Paper prototyping - what is it good for? *CHI '03 extended abstracts on Human factors in computing systems - CHI '03*, page 778.

Stoffregen, T. A., Bardy, B. G., Smart, L. J., and Pagulayan, R. J. (2003). *On the nature and evaluation of fidelity in virtual environments: Applications, Implications, and Human Performance*, pages 111–128. Lawrence Erlbaum Associates, Inc.

Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., and Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: Current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 521–530, New York, NY, USA. ACM.

Walker, M., Takayama, L., and Landay, J. A. (2002). High-Fidelity or Low-Fidelity, Paper or Computer? Choosing Attributes when Testing Web Prototypes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 46(5):661–665.

Zaman, B. and Shrimpton-Smith, T. (2006). The facereader: Measuring instant fun of use. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pages 457–460. ACM.

Zarour, M. and Alharbi, M. (2017). User experience framework that combines aspects , dimensions , and measurement methods. *Cogent Engineering*, 4(1):1–25.