Personal and Ubiquitous Computing (2023) 27:343–361
https://doi.org/10.1007/s00779-022-01696-8

ORIGINAL PAPER

Evaluation in scenarios of ubiquity of technology: a systematic literature review on interactive installations

Yusseli Lizeth Méndez Mendoza[1] · Emanuel Felipe Duarte[1] · Maria Jêsca Nobre de Queiroz[1] · M. Cecília C. Baranauskas[1]

Received: 1 August 2021 / Accepted: 18 October 2022 / Published online: 11 November 2022

Abstract
Literature in the field of human–computer interaction (HCI) has shown a long tradition of evaluation methods for and along with interactive systems design. We have experienced in the last years an impressive development in ubiquitous and pervasive systems, motivated by technological development, low cost of sensors and actuators, and a rise in the maker culture for the construction of computational systems. While such systems naturally inherit methods for evaluating the user interaction from previous interaction paradigms, it is not clear whether they reach specificities of the interaction of people within ubiquitous and pervasive systems scenarios. This work aimed at shedding light on this subject by conducting a systematic literature review on ubiquitous and pervasive technology scenarios of interactive installations. Results have shown that most of the selected contributions use classical methods of data collection and analysis, and combinations of these methods. Analysis of results also points out some new aspects to be considered in evaluation methods, regarding the human (social) actions promoted or afforded by ubiquitous and pervasive systems.

Keywords Evaluation methods · Ubiquitous systems · Interactive installations · Systematic literature review

1 Introduction

The evaluation of interactive systems is a core activity within the field of human–computer interaction (HCI). With its long tradition of evaluation methods, HCI literature has shown, for instance, well-tested methods to assess whether a user interface is easy to use, or to compare different designs in a decision-making process. As computational systems continuously evolve not only in terms of technology but also in interaction paradigms and their roles within human life and society, it follows that evaluation methods are expected to evolve too.

Computer use is now present in many areas of our lives and is used for a variety of purposes at all times, in all places, and by different people. These transformations are consequences of technological advancements such as the availability and affordability of sensors and actuators, as well as by a rise in the maker culture to the construction of computational systems. These changes, however, are not only of technological nature as they give rise to new forms of interactions. Since the seminal work of Mark Weiser [69], ubiquitous and pervasive technologies have created new forms of interaction, which expand to the physical environment.

Highlights
- Scenarios constructed with ubiquitous technology involve dimensions that might not be covered by current evaluation methods.
- A systematic literature review on the evaluation of interactive installations is conducted.
- Results raised the mainstream methods for data collection and analysis in interactive installations.
- The work reveals a complex relationship between categories of data collection and objects of evaluation.
- Missing aspects are related to the experience concept, the reachness of methods, and ethical issues.

✉ Yusseli Lizeth Méndez Mendoza
yusseli.mendoza@ic.unicamp.br

Emanuel Felipe Duarte
emanuel.duarte@ic.unicamp.br

Maria Jêsca Nobre de Queiroz
maria.queiroz@students.ic.unicamp.br

M. Cecília C. Baranauskas
mccb@unicamp.br

[1] Institute of Computing, University of Campinas (UNICAMP), Campinas, SP, Brazil

Weiser already foresaw technology "weaving into the fabric of everyday life" [69], p. 94] becoming indistinguishable from it, arguing that the technologies with the most profound effects are those that "disappear". For him, the idea of a personal computer, as well as the laptops emerging at the time (Dynabooks and "knowledge browsers"), represented only a step to reach the true potential of information technology. While these machines are the focus of attention, they do not really make an integral and invisible part of people's lives. Weiser's idea of ubiquitous computing goes beyond concerns with "the user interface," as he presented a new way of thinking about computers and taking the human world into account, allowing computers themselves to "disappear" into the background.

Adding to the concept of pervasiveness and ubiquity of computer systems, Kaipainen et al. [56] contributed to the idea of interaction as a human system coupling. Their concept of "enactive system," drawing on the seminal work of Varela et al. [66] on enactivism, involves the idea of considering human and computational not as separate systems, but as coupled parts of a single system. Instead of the standard explicit interface, interaction is described as embodied, i.e., it is based on unconscious psycho-physiological reactions by the body's involvement, and the human agent's spatial presence. Such a system presupposes that technology impacts directly on the human agent, whose experience impacts back on the technology. In practical terms, an enactive system uses sensors to detect both deliberate and non-deliberate information from the body (e.g., movement or physiological readings) and actuators to respond accordingly. This, in turn, generates a response in the person, and the enactive cycle continues.

The literature in the last decade, especially in the fields of HCI and Tangible User Interfaces (TUIs), has shown relevant research studies on scenarios of ubiquitous technology, conducted in laboratory e.g., [19] and in closed and public spaces (e.g., [36, 43, 65]). Although there is a wide range of contexts and focus in the studies, they are mostly interested in understanding the experience of people in these scenarios, their behavior, perception, connectedness, and engagement interacting with elements of the scenario. Contexts of museums, for example, have more recently created personalized sensorial experiences for visitors based on technologically extended interactive installations, constituting hybrid exhibition spaces, which bridge the material and physical with the technologically mediated and virtual [65]. Other authors, as contemporary artists and performance makers, have proposed the work with the esthetic experience of the body in computationally mediated installations to explore ways of imaging and experiencing the self and the world [19]. In our view, interactive installations represent good instances of ubiquitous and pervasive technology scenarios in their diversity and complexity, the reason why they are taken as objects of investigation in this work.

While these new ubiquitous, pervasive, enactive, systems may inherit some well-established traditional methods for evaluation of the user interaction, it is not clear whether these methods reach specificities that characterize the interaction of people within ubiquitous and pervasive system scenarios. Höök et al. [55], for example, raise concerns about directly applying HCI evaluation methodologies to interactive artworks, due to fundamental conflicts in worldview and methodology in this hybrid area. As discussed by Loke et al. [19], there is a tension between artists as design authors and the audience interpretation of the work; the evaluation data may be approached differently by artists and interaction designers, depending on the values underlying the design process. Also, some attributes of good interaction design, such as "engagement" seem to demand a fresh understanding when dealing with ubiquitous technology environments. Dalsgaard et al. [8], for example, show that the issue of engagement, which in the interaction design literature has focused on particular design strategies and attributes for promoting or supporting engagement, must be considered from a different perspective. For them, an approach is needed to capture engagement as a product of relations between physical, cultural, social, and content-related elements. In their study of an installation for embodied social active listening to sound and music content, Volpe and Camurri [43] point out the need of new social features which include leadership, saliency, synchronization, contagion, and co-creation. In particular, they mention the need for measuring the degree of emotional synchronization among the participants, for both the basic research perspectives and the implied applications.

Ultimately, new aspects of interaction in ubiquitous technology scenarios challenge the mainstream methods of HCI to the design of interactive systems, demanding new ways to consider interaction design and its evaluation, pushing related fields forward. Moreover, we wonder whether Weiser's dream of the disappearing computer is being effectively realized. Referring to interactive installations in museum contexts, for example, Torpus [65] raises the problem of the information layers brought by the technological gadgets in attracting the attention of visitors to the screens, competing with the other elements of the scenario, the exhibition content itself, and other people around.

Considering the scatterness of focus and methods in the literature, this work aims at shedding some light on this subject, by presenting and discussing the results of a systematic literature review on the evaluation of scenarios constituted by ubiquitous and pervasive technology. As an instance of ubiquitous and pervasive technology scenarios, our literature review examined the context of interactive installations. This specific context is appropriate for our investigation also because interactive installations and their exhibition spaces are often in the avant-garde of interaction design by their constant experimental use of technology and envision of novel, unconventional

Table 1 Research questions and motivations

Research question	Motivation
RQ1: How technology-based interactive installations have explored the concepts of ubiquitous, pervasive systems, and embodied cognition?	This question is important to identify how emergent theories and concepts are put into practice in the context of installations. To answer this question, it is necessary to analyze the concept and implementation of interactive installations in the selected literature
RQ2: Which methods and approaches are used to evaluate interactive installations in exhibition spaces?	This question is important to identify strengths and missed aspects in the methods used by the studies. To answer this question, it is necessary to identify what methods they used, what was evaluated and how, and who was involved
RQ3: Are evaluation methods taking into account the new types of (social) actions and perceptions afforded by interactive installations? If so, how?	This question is important to identify the social aspects involved in the evaluation conducted by the studies. To answer this question, it is necessary to consider the social affordances of the installations and information from the context of the studies

interaction approaches. With this systematic literature review, we aim at mapping the terrain to aspects that are being evaluated within scenarios of ubiquitous and pervasive technology constituted by interactive installations, evaluation methods which are applied, who is the target audience, and whether the evaluation methods are taking into account the new types of (social) actions and perceptions afforded by interactive installations. We also raise aspects deserving more investigation that have been missing in the literature. Thus, the paper contributions are twofold: (1) A map and a discussion of current evaluation practices, to help HCI researchers and practitioners choose their methods in designing or evaluating scenarios of ubiquitous and pervasive technology and (2) directions for further research to amend the shortcomings of current evaluation methods with new considerations.

This article is structured as follows: In Sect. 2, we present the methodology of our systematic literature review, including the steps that were conducted and method of analysis of the results; in Sect. 3, we present the results from the systematic literature review with an overview on how evaluation is treated by the selected works, including what is evaluated, how it is evaluated, and who participates in the evaluation; in Sect. 4, we provide answers to the review research questions, discussing main findings, and presenting some new aspects to be considered in methods for evaluating the human (social) actions promoted or afforded by ubiquitous and pervasive systems; lastly, in Sect. 5, we present our conclusions and point out directions for further work.

2 Review methodology

In this systematic literature review, the conducted process [52] started with defining the review team composed by the four authors of this article and establishing the objective of investigating evaluation practices in the universe of ubiquitous and/or pervasive technologies, instantiated in the context of interactive installations. The review process includes stages like specifying the research questions, establishing the review protocol with the formulation and conduction of a search strategy and selection criteria, describing the studies' characteristics, specifying the study quality and relevance assessment, and accomplishing the data synthesis, which are further analyzed and discussed in this work.

2.1 Research questions

The research questions that guided our search, and their motivations are presented in Table 1. They reflect an encompassing effort to understand how scenarios of ubiquitous and pervasive technologies instantiated in interactive installations have addressed evaluation in these scenarios.

2.2 Review protocol

The review protocol was based on the PRISMA [61] flow diagram, as illustrated by Fig. 1. The flow diagram maps out the number of records identified, screened, included, and excluded, and the reasons for exclusions. The main components of the review protocol are properly described in the subsequent subsections.

Search strategy We selected four digital libraries to search for research studies: ACM Digital Library[1] with the search expanded to also include the larger database known as "ACM

[1] https://dl.acm.org/

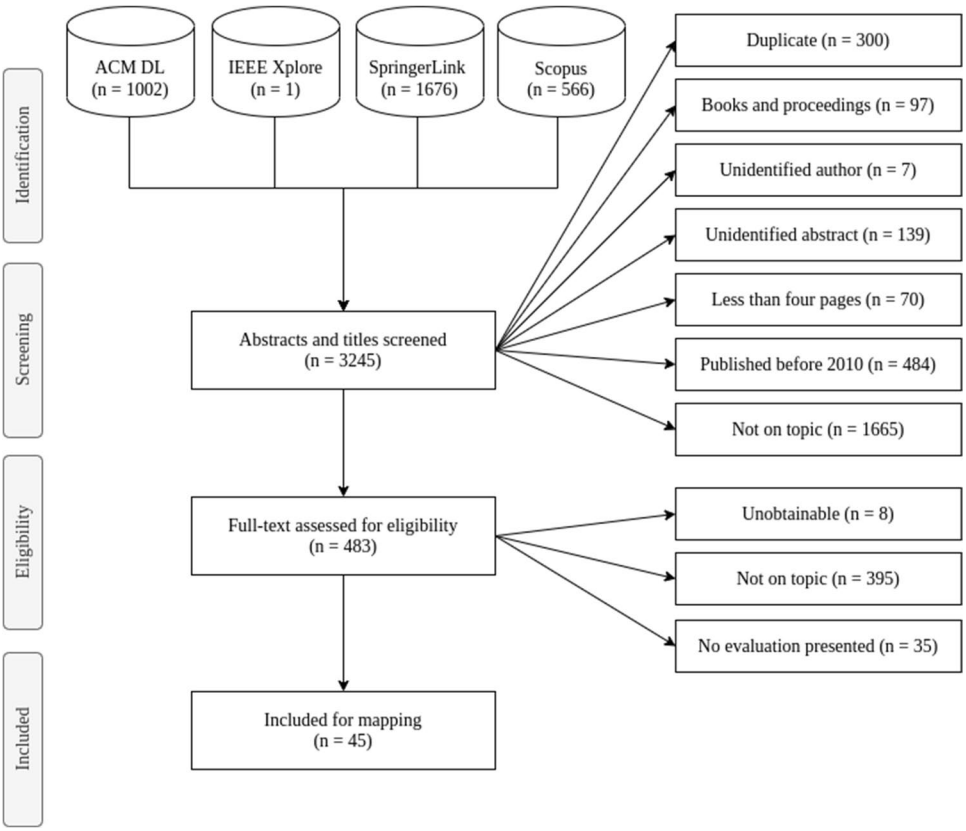

Fig. 1 Search and selection flow diagram. Based on the PRISMA Flow Diagram [61]

Guide to Computing Literature", IEEE Xplore,[2]SpringerLink,[3] and Scopus.[4] We chose these digital libraries due to their relevance in the field of Computer Science and Human–Computer Interaction (HCI). We also considered the Scielo[5] digital library because of its overall relevance in Latin America, although it did not return any results to our query.

Based on our research questions, we devised a search string composed of four main parts. First, the search string should screen for at least one term from a list of terms related to ubiquitous, pervasive, enactive, sentient, embodied, and embedded. Second, to investigate how these concepts have been applied in the literature, the string should screen for documents that present some kind of system or environment with some type of technology, likely digital technology. Third, the string narrows the search to our desired specific context of interactive installations with the terms: interactive installation, art installation, installation art, multimodal installation, and museum exhibition. Lastly, the string should screen for documents with the terms enact, action, and perception, with the objective of addressing embodied forms of interaction. It is also worth noting that the search string does not use any keyword related to evaluation because we aimed at reaching a wider universe of studies, regardless of whether they feature evaluation or not. The search string was written in the following format:

(ubiquitous OR pervasive OR enactive OR sentient OR embodied OR embedded).

AND (environment OR system) AND (technology OR digital).

AND ("interactive installation" OR "art installation" OR "installation art" OR "multimodal installation" OR "museum exhibition").

AND (enact OR action OR perception).

Inclusion and exclusion criteria In order to select the most appropriate studies for our investigation, a set of inclusion and exclusion criteria was defined as illustrated by Table 2.

After the elimination of duplicate studies (EC1), exclusion criteria EC2 and EC3 are necessary because we are not able to properly assess studies that do not indicate the authors or provide an abstract. Exclusion criterion EC4 considers that studies with less than four pages are unlikely to contain sufficient material to contribute towards answering our research questions. Exclusion criterion EC5 considers that studies published after 2010 are more likely to contribute due to the state-of-the-art of the systems in technological terms. Exclusion criterion

[2] https://ieeexplore.ieee.org/

[3] https://link.springer.com/

[4] https://www.scopus.com/

[5] https://scielo.org/

Table 2 Inclusion and exclusion criteria

Exclusion criteria	Inclusion criteria
EC1: Duplicate studies	IC1: The document features interactive installations as a central aspect
EC2: Studies with unidentified author(s)	IC2: The document features some form of evaluation in the study
EC3: Studies without abstract	IC3: The document features social aspects and/or interactions in the presented system
EC4: Short studies (less than four pages)	IC4: The document features embodied cognition and/or related concepts
EC5: Studies published before 2010	
EC6: Studies not written in English, Portuguese, or Spanish	
EC7: Full books or proceedings	
EC8: Studies that diverge from the subjects of the research questions	

EC6 excludes studies not written in English, Portuguese, or Spanish. This criterion takes into account that English is the predominant scientific publication language within Computer Science and HCI. Moreover, we also consider studies written in Portuguese and Spanish because these are the authors' native languages. Exclusion criterion EC7 states our interest in studies published as journal articles, papers from conference proceedings, and book chapters. Lastly, the exclusion criterion EC8 is tied directly to studies with the potential of answering our research questions.

Otherwise, the inclusion of studies was determined by the criteria shown in Table 2. Inclusion criteria IC1, IC2, IC3, and IC4 address specific topics of interest in our systematic literature review: interactive installations in exhibition spaces (IC1), some form of evaluation (IC2), social aspects and/or interactions (IC3), and embodied cognition and related concepts (e.g., enaction, embodiment, coupled action, and perception) (IC4). For a study to be selected, it should satisfy IC1 and IC2 and at least one between IC3 and IC4. The necessity of IC4 is justified by the importance of the concept of embodied cognition and related concepts for ubiquitous and pervasive systems, and IC2 is justified by our focus on evaluation in this article. For IC3 and IC4, while individually important, we considered that requiring the selection of both of them was too restrictive.

Excluded = EC1 OR EC2 OR EC3 OR EC4 OR EC5 OR EC6 OR EC7 OR EC8
Included = NOT Excluded AND IC1 AND IC2 AND (IC3 OR IC4)

Search and screening After the formulation of the search strategy and the selection criteria, we proceeded to submit our search string to the selected digital libraries, using the full-text search. The string syntax was adjusted according to the parameters of each digital library when needed while preserving its logic. The search was performed on July 25, 2020. Initially, 3245 studies were identified through database searching: the ACM Digital Library returned 1002 results, IEEE Xplore returned 1 result, SpringerLink returned 1676 results, and Scopus returned 566 results. This phase is labeled as "identification" in the flow diagram of Fig. 1. The retrieved records were exported in BibTeX format, and we used the JabRef reference management software to normalize the records to be screened according to our selection criteria.

In an initial "screening phase," we excluded 300 duplicate entries (EC1), 7 entries with unidentified authors (EC2), 139 entries without abstract (EC3), 70 entries with less than four pages (EC4), 484 entries published before the year of 2010 (EC5), and 97 entries for full books or proceedings (EC7). No entries satisfied exclusion criterion EC6. To continue the screening phase, we exported the JabRef records to a shared spreadsheet for the manual screening of titles and abstracts by the authors. An amount of 1665 studies were deemed not on topic, i.e., showed no clues towards contributing to our research questions (EC8), and therefore were also excluded. Thus, a total of 483 studies were included to be assessed for eligibility. We downloaded the full text for these potentially relevant studies, except for 8 entries to which we were not able to obtain the full text. We skimmed through all the entries' full-texts paying attention to inclusion criteria IC1, IC2, IC3, and IC4. A total of 395 entries were deemed not on the topic, 169 of which satisfied exclusion criterion EC1. The other 226 entries did not meet our rule of inclusion by satisfying both inclusion criterion IC1 and at least one among IC3 and IC4. Additionally, a total of 35 entries were excluded because they did not present some form of evaluation (IC2). Finally, a total of 45 studies remained for final review and were included for analysis.

Table 3 Quality parameters

Quality parameter	Scoring
QP1: Is the methodology of the study described in detail?	Yes = 1; no = 0
QP2: Is a design process described in detail?	Yes = 1; no = 0
QP3: Is there any user participation during the evaluation?	Yes = 1; no = 0
QP4: What is the percentile of citations per year of the document?	Max = 1; min = 0
QP5: What is the h5-index percentile of the journal or conference?	Max = 1; min = 0

2.3 Description of study characteristics for the selected studies

We thoroughly read each of the 45 selected full texts and proceeded to fill in a form with questions aimed at mapping and describing the characteristics of the studies. The form asked direct questions such as what kind of institutions are the authors affiliated to, and where are they from; whether there is practical application in the study, and if so, what is the application context and the target audience; what is the methodology of the study; and whether the study has some kind of evaluation, and if so, what is evaluated and how. We were mainly interested in questions related to the evaluation proposed or conducted by the studies. For example, in the question: "Does the study have some kind of evaluation?" If so, "what is evaluated?" The form provided us with a set of evaluation topics to be selected as check-boxes such as (a) artifact and design, (b) people's experience, (c) scenario, (d) interaction, and (d) other, with open edit text. For the question related to the evaluation methods and/or tools applied in the study, the form provided us with a set of evaluation methods also to be selected as check-boxes such as (a) video analysis, (b) automatic analysis, (c) questionnaire, (d) interview, and (e) other, with open edit text. In both cases, the "other" option had an open field for additional comments and suggestions. This set of questions contributed to answers to research questions 2 and 3.

Among other topics related to the technology used, the types of interaction, and the social aspect involved in the works, the form asked questions such as "What computational technologies besides computers are used?", "How does the interaction with computational technology take place?", "How is the human body considered and involved in the interaction?", "How is embodied cognition considered and involved in the interaction?", "What kind of installation is portrayed in the study?", and "How are social aspects and interactions present in the study?" The data obtained from this set of questions contributed to answer to research question 1. We saved the form responses into a shared spreadsheet for data organization and analysis. The content of the spreadsheet included some of the following information: (a) study title, (b) author(s), (c) country(ies), (d) application context, (e) methodology, (f) target audience, (g) evaluation object, and (h) evaluation method. The analysis involved data accounting, grouping data into assigned categories, and triangulation of results (e.g., the association between data collection methods and evaluation subjects).

2.4 Quality and relevance assessment for the selected studies

To estimate the quality and relevance of the selected studies, we used two sets of parameters: quality parameters and relevance parameters. The quality parameters can be considered as more objective, general evidence of the quality of the studies selection, whereas the relevance parameters can be considered as general evidence of the relevance of the studies towards our research questions. The quality parameters are presented in Table 3, while the relevance parameters are presented in Table 4.

The quality parameters QP1 to QP3 have a binary response, if the answer is "yes," the value of the parameter will be 1; otherwise, it will be 0. The responses for these parameters are determined by our thorough reading of the full texts. The QP4 and QP5 parameters are derived from the citation and h5-index values, respectively, obtained from Google Scholar on December 7, 2020. For the QP4 quality parameter, we obtained the number of citations for each document, normalized the value by calculating the number of citations per year since the publication of the document, and calculated the percentile of each value of citations per year within our set of selected entries. For the quality parameter QP5, we first assigned the value 0 to studies with no h5-index value available; then, we calculated the h5-index percentile of the documents in our set of selected entries. After dividing each percentile by 100, both quality parameters QP4 and QP5 were left with answers ranging from 0 to 1 with two decimal places. Finally, the overall quality was calculated as a simple average of the values of the quality parameters QP1 to QP5.

For each relevance parameter from RP1 to RP6, the numerical value is calculated according to how many descriptors (from the form) were selected for that category. If no descriptor is selected, the value will be 0, if the number of selected descriptors is greater than 0, but

Table 4 Relevance parameters

Relevance parameter	Scoring
RP1: How much is the "technology" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0
RP2: How much is the "interaction" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0
RP3: How much is the "embodiment" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0
RP4: How much is the "enactive" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0)
RP5: How much is the "social" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0
RP6: How much is the "interactive installation" category covered?	Equal or above the median = 1 Below the median = 0.5 Not covered = 0

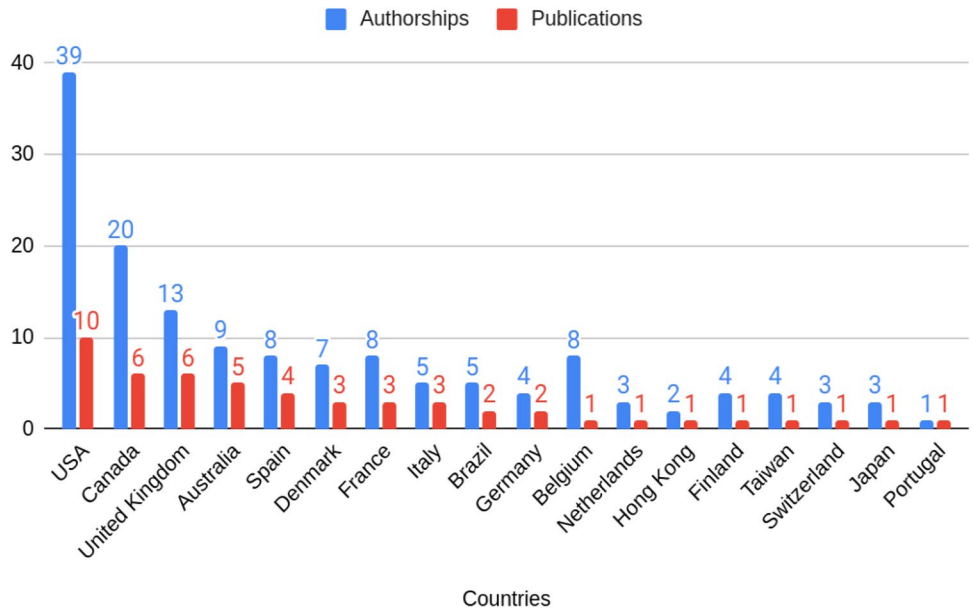

Fig. 2 Chart of authorships and number of publications by country

less than the median of the selected descriptors for the whole set, the value will be 0.5, and if the value is equal or greater than the median, it will be 1. To calculate a value of general relevance, we first added a weight of 3 for RP1 and RP2, and a weight of 5 for RP3 to RP6 because we considered these last four parameters to be significantly more important for our research questions. Then, we calculated a simple weighted average of the values.

$$Quality = (QP1 + QP2 + QP3 + QP4 + QP5)/5$$
$$Relevance = ((RP1 \times 3) + (RP2 \times 3) + (RP3 \times 5) + (RP4 \times 5) + (RP5 \times 5) + (RP6 \times 5))/26$$
$$Quality\ \&\ Relevance = (Quality + Relevance)/2$$

3 Results overview

A total of 45 studies were selected, featuring authors' affiliations distributed between 18 countries (we quantified all authorships of each study, e.g., a study with three authors has three authorships). Figure 2 shows the number of publications by country. The USA published the highest number of studies (10 documents) with 39 authorships, followed by Canada, UK, and Australia. Regarding the studies' research approaches, 26 out of 45 (57.78%) are qualitative research, 6 (13.33%) are quantitative research, and 9 (20%) combined both qualitative and quantitative approaches. In 4 studies (8.89%), we were not able to clearly identify a

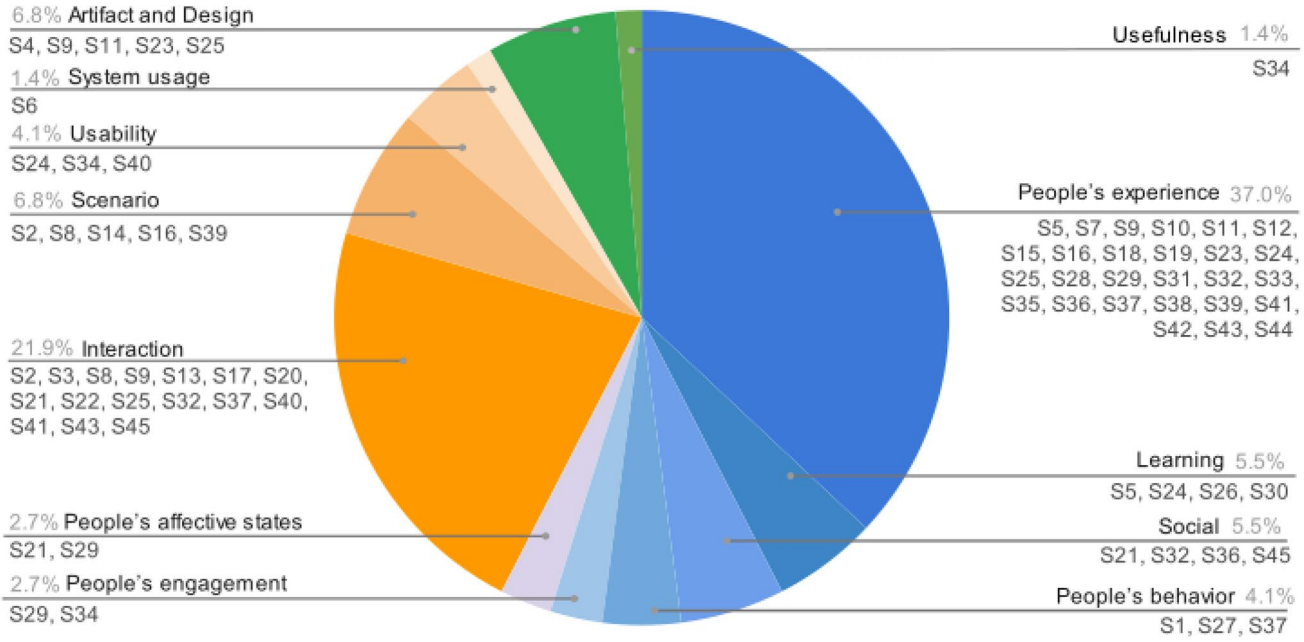

Fig. 3 Distribution of categories of evaluation

specific methodological approach. Case studies were the most applied research method of the qualitative approach.

The selected studies were published in 34 different journals, conference proceedings, and books. Among them, we highlight that there were 4 papers published in the ACM Conference on Human Factors in Computing Systems (CHI), 4 papers published in the Audio Mostly (AM) conference, 3 papers published in the Australian Conference on Computer–Human Interaction (OzCHI), 2 papers published in the Brazilian Symposium on Human Factors in Computing Systems (IHC), 2 papers published in the Conference on Creativity and Cognition (C&C), and 2 articles published in the Cognitive Computation journal. With the exception of this last journal, which is published in SpringerLink, the other conferences are all published in the ACM Digital Library. Besides these highlighted vehicles, the remaining 28 journals, conference proceedings, or books, each contains only a single publication.

The next subsections organize the resulting data on the subject of evaluation (i.e., what is being evaluated), the audience and people's involvement (i.e., who and how many people have interacted with the computational system in question), and the methods applied (i.e., how evaluation was conducted by the researchers).

3.1 The categories of evaluation

We identified 12 categories of evaluation: 6 of them target people directly: people's experience, learning, social, people's behavior, people's engagement, and people's affective states; 2 studies address system concerns such as usefulness, artifact and design; lastly, there are 4 categories that are "in between," focusing on both people and system: scenario, interaction, usability, and system usage. This categorization is based on the set of evaluation topics defined in our form as detailed in Sect. 2.3. An overview of the relative distribution of evaluation categories among the studies can be seen in Fig. 3.

Results in Fig. 3 reveal that *people's experience* was the main aspect evaluated by researchers with 27 occurrences. This category involves the evaluation of overall experiences (e.g., [29, 38]); works interested in evaluating the user experiences (UX) (e.g., [24, 41, 44]) by applying a variety of UX instruments; and other studies interested in evaluating bodily experiences designed by applying somatic connoisseurship (e.g., [10, 19, 28, 33]). Some works considered the engagement and affective states of people in the evaluation of their experience (e.g., [9, 12, 32, 44]); thus, these works were grouped in this category. Works that evaluate functional properties of an artifact or the user perception of these properties in the evaluation of the people's experience (e.g., [41]) were also categorized as such. The *interaction* was another main aspect evaluated by researchers with 16 occurrences, for example, by observing participants interacting with an artifact to identify interaction patterns (e.g., [13, 43]). The other categories are listed as follows:

The *artifact and design* category, with 5 occurrences, refers to the evaluation of an artifact in terms of its design characteristics by using, for example, guidelines [4], models [9, 11], or video recordings [23, 25].

The *scenario* category, with 5 occurrences, includes works that take into account the environment in which the participants are immersed to evaluate an artifact or system [2, 8, 14, 16, 39].

The *learning* category, with 4 occurrences, refers to the evaluation of learning environments developed to facilitate the teaching, learning, and understanding of different concepts using, for example, tangible artifacts [24] or the human body in the interaction [5, 26, 30].

The *social* category, with 4 occurrences, refers to the evaluation of an artifact mainly focusing on social presence and collaboration [21, 32, 36, 45].

The *people's behavior* category, with 3 occurrences, refers to the evaluation of people's behavior when interacting with systems designed to influence or change their behavior, for example, by applying persuasive design (e.g., [37]).

The *usability* category, with 3 occurrences, refers to the perceived system's usability by people [24, 34, 40].

The *people's engagement* category, with 2 occurrences, refers to the evaluation of the capacity of an artifact or system to motivate people to interact with technology. The studies included in this category [29, 34] used specific instruments to collect data about the engagement.

The *people's affective states* category, with 2 occurrences, involves studies interested in designing and evaluating systems aimed at changing people's affective states [21, 29].

Lastly, *system usage* and *usefulness* were the categories least considered in the evaluation by researchers, with 1 occurrence for each case [6, 34].

3.2 The audience and people's involvement

Regarding the audience and people's involvement in the evaluation, we considered the number of people reported in the studies; people who interacted with an artifact just as visitors of an installation, for example, were not considered. Regarding sample size, we found 24 studies with less than 50 participants, 3 studies with a sample between 50 and 100 participants, 5 studies that involved a larger number of participants (more than 100 people), and 13 studies that did not report information about the number of participants. Considering the works that mention the target audience of the study or that it was possible to infer their audience from the authors' description (e.g., professionals or college students are expected to be adults), in 21 studies, the target audience was of adults [5, 7, 9, 10, 12–15, 17–20, 24, 28, 29, 32, 33, 37, 38, 42, 44]; in 17 studies, the considered target audience was of children, adolescents, and adults [1–4, 6, 8, 22, 23, 27, 30, 34–36, 40, 41, 43, 45]; in 3 studies, the target audience was children [21, 25, 26]; and 1 study was aimed at children and adolescents [16]. This distinction of age groups is relevant when considering that children and adults may require different evaluation approaches and methods.

Regarding the study context, 21 of the 45 studies refer to exhibition contexts with 15 being in public spaces, such as museums and urban areas [1, 2, 4, 8, 23, 26, 27, 30, 31, 35, 37, 39, 40, 44, 45] and 6 being in closed spaces, such as academic conferences and closed exhibitions [3, 6, 7, 36, 41, 43]. The other 24 studies were evaluated in laboratory contexts (i.e., a controlled environment) [5, 9–22, 24, 25, 28, 29, 32–34, 38, 42]. Furthermore, we found that 11 out of the 45 studies explicitly mention either approval from an ethics committee or at least informed consent from the participants [2, 12, 16, 19, 23, 24, 26, 29–31, 45]. In exhibition contexts, there is no control of who may interact with an installation and how, while in laboratory context participants can be selected and follow a protocol.

3.3 Data collection methods

We identified six categories of data collection methods in the evaluation processes of the studies: (1) interview, (2) video recording, (3) questionnaire, (4) observation, (5) system logs, and (6) physiological measurement. An overview of the distribution of these categories is illustrated in Fig. 4. It is worth noting that some works use more than one data source for evaluation. Details about each category are reported in what follows.

Interview The most used data collection method was the interview, with 21 occurrences. Results revealed structured and semi-structured interviews (e.g., [5, 10, 12, 13, 15, 17, 18, 29, 37, 38, 40, 41]), Video-Cued Recalls (VCR) (e.g., [5, 31]), and second-person interview (e.g., [10]). A structured interview presents questions in a well-defined order and there is no room for adding non-predefined questions, whereas with a semi-structured interview it is possible to gain additional insights and understanding by asking interviewees for clarification, following their comments, or adding new questions [58].

The *Video-Cued Recall (VCR)* interview method [60] employed video support to evaluate embodied experiences reducing the risk of self-bias or self-reporting that can be present in surveys and other interview methods without video support. For instance, in the [5] study, participants were asked open-ended questions about their experience as they watched the video of their dancing session. The *second-person interview* or *explicitation interview* [63] is a method for collecting the first-person perspective of interviewees about their experience. This method was developed from the seminal work of Pierre Vermersch on explicitation [67]. The main goal of this method is to facilitate the re-living of a specific past experience in greater detail so that it can be lived again in the present. For instance, in the [10] study, the interviewer selected singular experiences, e.g., particular gestures or sensations, from the observation

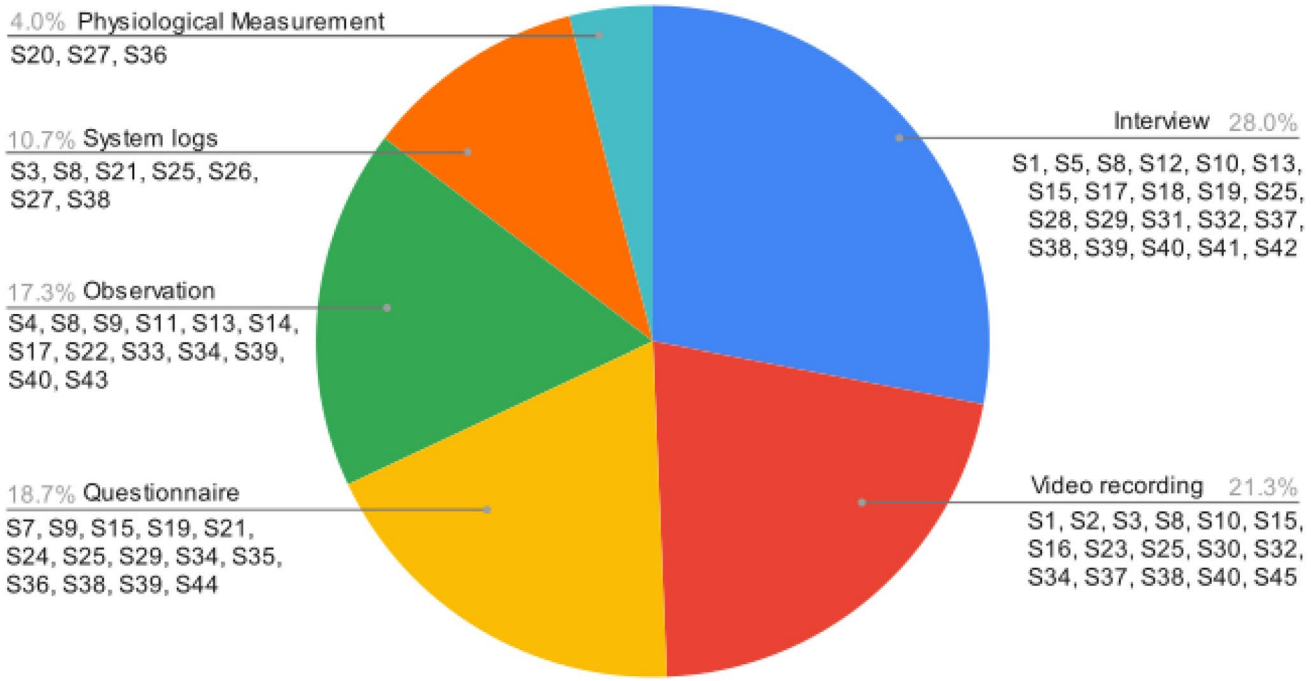

Fig. 4 Distribution of categories of data collection methods

of the participant's interaction. Then, the interviewer collected detailed descriptions from the participants on these singular experiences. The interviewer was responsible for guiding the participant's attention to previously unnoticed or forgotten aspects of these moments.

Video recording A total of 16 works [1–3, 8, 10, 15, 16, 23, 25, 30, 32, 34, 37, 38, 40, 45] collected data using video recordings to capture, for example, observational bodily movements (e.g., [37]), interactions with technology, and the physical space where participants were located (e.g., [2, 8, 30]). Some works used audio recordings as a complement to video recordings (e.g., [10, 40]).

Questionnaire This category grouped data collection methods involving data collected from questionnaires and surveys, with 14 occurrences: [7, 9, 15, 19, 21, 24, 25, 29, 34–36, 38, 39, 44]. Overall, questionnaires included open-ended questions, rating scale questions, or Likert scale questions. Also, we found studies using instruments based on already existing questionnaires. For instance, the [44] study used a UX questionnaire split into four parts, including a positive/negative affect (PANAS) scale, word pairs of the AttrakDiff questionnaire, and open-ended questions.

The PANAS scale is one of the most widely used scales to measure mood or emotion [68]. This scale is composed of 20 items, with 10 items measuring positive affect (e.g., excited) and 10 items measuring negative affect (e.g., scared). Each item is rated on a five-point scale. The AttrakDiff questionnaire, in turn, is one of the most widely used instruments to evaluate the user experience with interactive systems [50, 54]. This questionnaire is composed of 28 semantic differential questions presented as seven-point scales. The two ends of the scale host opposite adjectives (e.g., complicated – simple, ugly – attractive).

In the [44] study, researchers included only 9 word pairs of the original AttrakDiff. Another work that elaborated an instrument based on already existing questionnaires was the [21] study. The questionnaire included rating scale questions, questions based on a five-level Likert scale, and a section with an adapted version of the Self-Assessment Manikin (SAM) scale [46]. This last scale is a pictographic scale to measure emotion using the Pleasure, Arousal, and Dominance (PAD) emotional state model. Each item can be assessed on a 5- or a 9-point scale. In [21], researchers used the 9 points scales for pleasure and arousal dimensions.

In the [35] study, we found a questionnaire that included open-ended questions, questions in Likert scale format, and a section based on the *PLEX Cards* [59]. The PLEX Cards can be applied to design and evaluate playful experiences and were created to communicate each of the 22 PLEX framework categories [57] (e.g., captivation, competition, discovery). The PLEX framework is an expanded version of the pleasure framework [49]. In [35], the authors used a set of 12 cards and they presented only one descriptive picture, instead of the original two.

Regarding the evaluation of people's engagement, we found two types of questionnaires: in the [29] study, the authors used the *game engagement questionnaire* [48] that measures levels

of engagement by considering four categories: flow, immersion, engagement, and presence; whereas in the [34] study, authors used the *O'Brien's Engagement Questionnaire* [62] by considering six attributes: perceived usability, esthetics, focused attention, felt involvement, novelty, and endurability. As for the evaluation of social aspects of interaction, in the [36] study, the authors used the *Condensed Networked Minds Social Presence Inventory (NMSPI)* [53] to measure levels of social presence, whereas, in the [21] study, the authors used a questionnaire to measure levels of perceived collaboration and impressions of group members.

Observation A total of 13 works collected data through note taking of interaction situations [4, 8, 9, 11, 13, 14, 17, 22, 33, 34, 39, 40, 43]. For instance, in [9] the participants were observed in terms of where they looked at and which reactions the installation produced upon their actions. The observations were captured with a standardized form, according to the authors. In [8], researchers collected data through in situ observations supplemented by video recordings of use situations, whereas in [4] the observations were based on proposed guidelines. In some works, the observations allowed researchers to raise a diverse set of interaction patterns for classification (e.g., [13, 43]).

System logs Results revealed that, in 8 works, researchers saved data in log files [3, 6, 8, 21, 25–27, 38]. These log files stored quantitative data generated by the system itself such as the duration and number of registered interactions (e.g., [6, 8, 38]); whereas in technological environments that incorporate multi-touch surfaces or capture full-body interactions, researchers used these log files to record interaction positions on the interactive surface, or positions and rotations of the user's body when interacting with the system (e.g., [3, 21, 25, 26]).

Physiological measurement The interaction between people and their environment can be enabled through physiological data (e.g., Pulse, Breath, Heart Rate) that can generate changes in their environment. We identified three studies in this category [20, 27, 36]. For instance, the [27] study presented a public installation in which participants controlled the environment lights using their heart rate, and the [36] study presented a virtual reality installation for exploring breathing synchronization and during the evaluation, data from the breathing sensor was collected.

3.4 Data analysis methods

Qualitative data such as interview transcripts, written notes, and video recordings, compose the data collected for analysis. These collected data were analyzed using a variety of analytic methods such as thematic analysis (e.g., [12, 13, 45]), grounded theory (e.g., [10, 29]), and interpretative phenomenological analysis (IPA) (e.g., [5, 31]). *Thematic analysis* [47] is a qualitative method for identifying, analyzing, and reporting themes (patterns) within data. A theme captures something important about the data in relation to the research question and represents some meaning within the data. The thematic analysis process included transcription, coding, the definition of themes, and a written report. *Grounded theory* [58] is a qualitative analysis method to generate a theory of the phenomenon that is grounded in the data gathered and analyzed. The method starts from a set of empirical observations or data and aims to develop a theory grounded from the data. The process includes transcription, coding, coding categories, and the formation of a theory. *The interpretative phenomenological analysis (IPA)* [64] is another qualitative analysis method that explores in detail how participants are making sense of their personal and social world.

Quantitative data were collected through questionnaires, system logs, and physiological measurements. Descriptive statistics (e.g., mean, median, standard deviation) was used from the data provided in questionnaires (e.g., [21, 24, 25, 34]). Some works used a Mann–Whitney U test (e.g., [24]) and others the F-test (e.g., [25, 29]) for data analysis. Regarding system logs, descriptive statistics were also applied on the data retrieved from log files such as information about how long the interaction lasted (e.g., [38]). Regarding physiological data, authors reported their results by comparing the values of the collected data with physiological thresholds such as the heart rate threshold (e.g., [27]).

3.5 Results of the quality and relevance assessment

The evaluation of the quality and relevance of the selected documents took into account 11 variables, forming two sets of metrics. The first set contains 5 variables that are related to more general aspects of the quality of the selected works (e.g., citation and h5-index metrics), while the second set contains 6 variables related to the relevance of the works for this investigation (e.g., the research questions).

The score calculated for the quality of the documents had an average value of 0.50 with a standard deviation of 0.10. The score calculated for the relevance of the documents, in turn, had an average value of 0.86 with a standard deviation of 0.17. Using these two values to calculate a combined score of quality and relevance, from a value ranging from 0 to 1, the 45 selected entries achieved a combined average score of quality and relevance of 0.68 with a standard deviation of 0.09. Of the 45 documents, 22 works scored between 0.84 and 0.70, 21 works scored between 0.69 and 0.50, 2 works scored less than 0.50, with these two scores being 0.48 and 0.36. All quality and relevance assessment scores

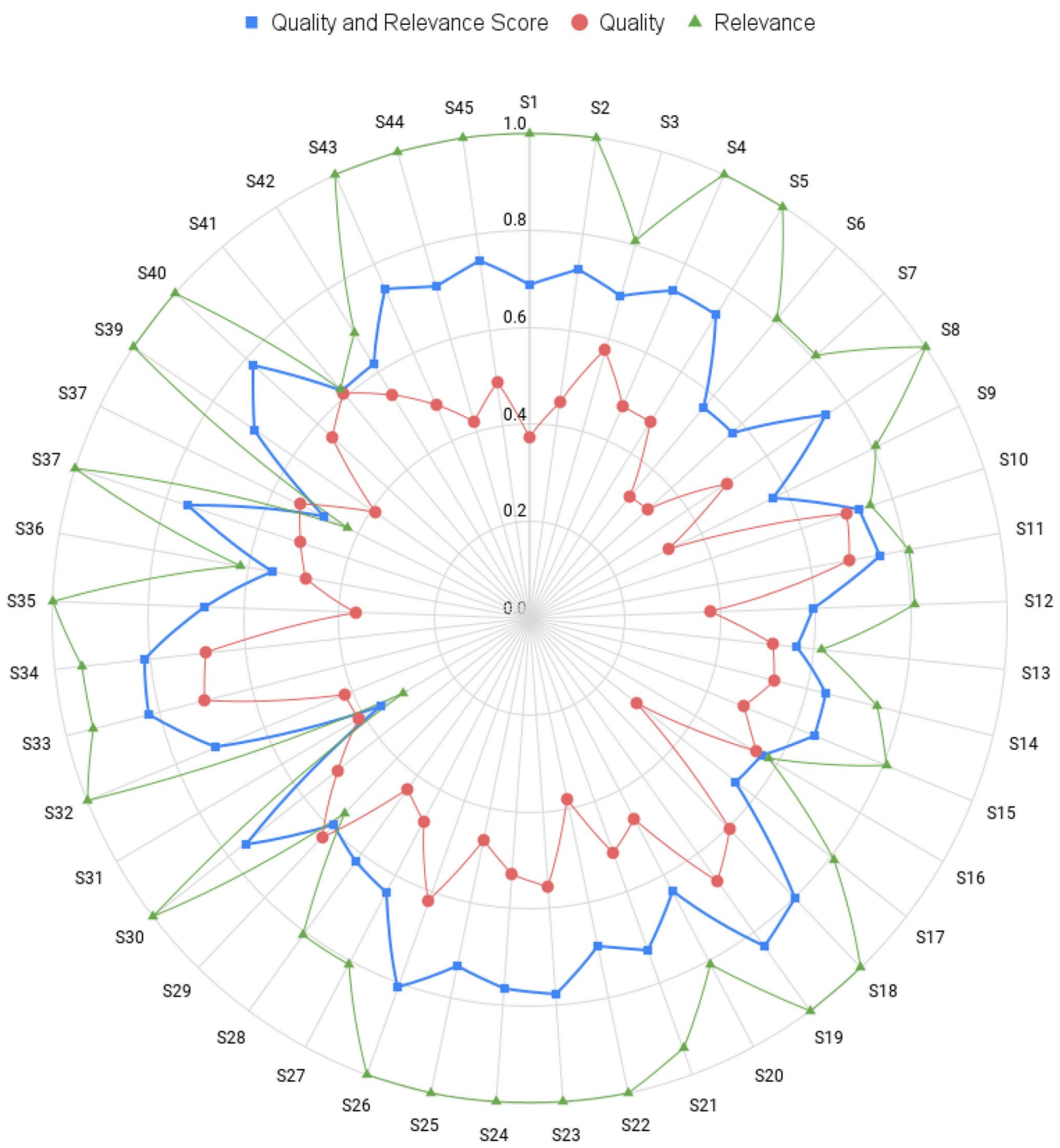

Fig. 5 Chart of the calculated score of the quality and relevance assessment

can be viewed in the graph shown in Fig. 5, which contains the individual values for quality and relevance scores, as well as the combined quality and relevance metrics. As shown in Fig. 5, the studies with the highest quality ratings are [10] (0.70), [33] (0.70), [11] (0.68), [34] (0.68), and [19] (0.67). Regarding the relevance of the studies, 21 obtained a maximum score with a score equal to 1.00, these being [1, 2, 4, 5, 8, 18, 19, 22–26, 30, 32, 35, 37, 39, 40, 43–45]. Regarding the combined score of quality and relevance, the studies [19] (0.84), [33] (0.82), [26] (0.81), [34] (0.81), and [18] (0.8) obtained the highest scores.

In the next section, we show how this corpus of works in the field leads us to answers for the research questions and to the discussion on issues still in need of deepening.

4 Synthesis and discussion

Based on results presented in Sect. 3, in this section, we answer the research questions presented in Table 1 (Sect. 2.1) identifying shortcomings in current evaluation practices. Then, we propose further topics for a research agenda on the field.

4.1 Answering the research questions

For RQ1, "How technology-based interactive installations have explored the concepts of ubiquitous, pervasive systems, and embodied cognition?" Most of the selected studies explicitly address the use of displays (22 occurrences), sensors (21 occurrences), microcontrollers (10 occurrences), and other embedded devices with wireless communication. For example, we found studies using displays and Kinect sensors (e.g., [13, 17, 22]), microcontrollers, and different types of sensors such as proximity or pressure sensors (e.g., [2, 4, 37, 40]). The artifacts are mainly related to tangible (e.g., [23, 34, 40, 45]), wearable (e.g., [28, 33, 35]), and robotic (e.g., [4, 7, 12]) technologies. The main types of interaction with the proposed artifacts illustrate mainly embodied and full-body interaction, mostly motion and gesture-based interaction (16 and 10 occurrences respectively) (e.g., [3, 4, 10, 26, 44]). In terms of underlying technology, we observe that most of the environments intend to address the main ideas originally proposed for ubiquitous, pervasive systems [69] as they tend to disappear in the background, into the periphery of our attention. Regarding underlying concepts and theoretical foundations, the works reveal concepts related to action and perception (19 occurrences each), sense-making (11 occurrences), autopoiesis (3 occurrences), and ontogenetic drift (1 occurrence), suggesting an embodied mind and embodied cognition approaches to understand the phenomenon of interaction in these environments (e.g., [3, 10, 17, 19, 44]).

Although the selected works seem to be conceptually near Weiser's idea of ubiquitous computing as "embodied virtuality," some drawbacks are still mentioned. In the context of museums, for example, the mode of interaction should not divert visitors from the exhibition's thematic and consequently take them out of immersion in that situation. Torpus [40] addresses the challenge of expanding museum exhibits through embedded media content without drawing visitors' attention to the invisible interface. Even using embedded technology, the attention of visitors still turns to the artifacts, and technology becomes an element that competes with the environment. Also facing problems related to the involvement of participants with autonomous interactive artworks, such as superficial levels of involvement and attention, Loke et al. [19] used a live art model, in which the visitor's experience was guided by an artist and technology was kept in the background while experience and body awareness gained prominence.

Still considering Weiser's approach to ubiquitous computing, another important aspect is the social; technology should allow users to be in the world; thus, interactions could be more social, allowing people to create connections. We observed, however, that in some interactions that involved body movements with feedback through displays, for example, the social aspect was not prioritized or properly taken into account in the design of the installations, with the experiences focused on the individual himself [3, 17]. In [17], for example, the user interacts with a life-size projection of himself and other bodies that have been tracked by a sensor, the installation combines mirrored video captured in real-time with recorded fragments of the past that resemble current movement. In summary, ubiquity is not only about availability or embeddedness of technology, but also about becoming an indistinguishable part of the environment, considering the people situated in it. It follows that as good practices, interactive installations in museums should not draw too much attention to their interaction or technologies, and these interactions and technologies should not hinder the social interactions that take place in such environments.

As for RQ2, "Which methods and approaches are used to evaluate interactive installations in exhibition spaces?" To address this question, we first observed *what* the works have considered relevant to evaluate, to discuss this relevance in terms of ubiquitous and pervasive systems values, as originally proposed. As presented in Sect. 3.1, the objects of evaluation in the 45 studies can be synthesized in three main groups: (1) the *what* involves people directly in their experience, behavior, learning, engagement, affective states, and social relations; (2) the *what* is system-focused in its usefulness, and design; and (3) the *what* is a "in between" human-technology relation, encompassing interaction, usability, usage, and overall scenario. From these, while some objects of evaluation seem to be inherited from conventional computational systems evaluation (e.g., usability, usage, usefulness, engagement), some new interests seem to stand out (e.g., affective states, and social aspects). Also, the whole scenario as an object of evaluation is consistent with the need to understand interaction in ubiquitous environments, in which its digital, material, technological, and social aspects may become increasingly interconnected and inseparable from one another.

Figure 6 illustrates the relations among categories of data collection methods (left) and *what* is being evaluated (right). The width of the associations represents the reach and frequency of data collection methods used. For instance, *interviews* and *video recordings* were the most used procedures to collect information about the experience, behavior, people's learning, and their interactions, including the interaction scenario, and also the artifacts and their design. Additionally, video recordings allowed researchers to assess usability and

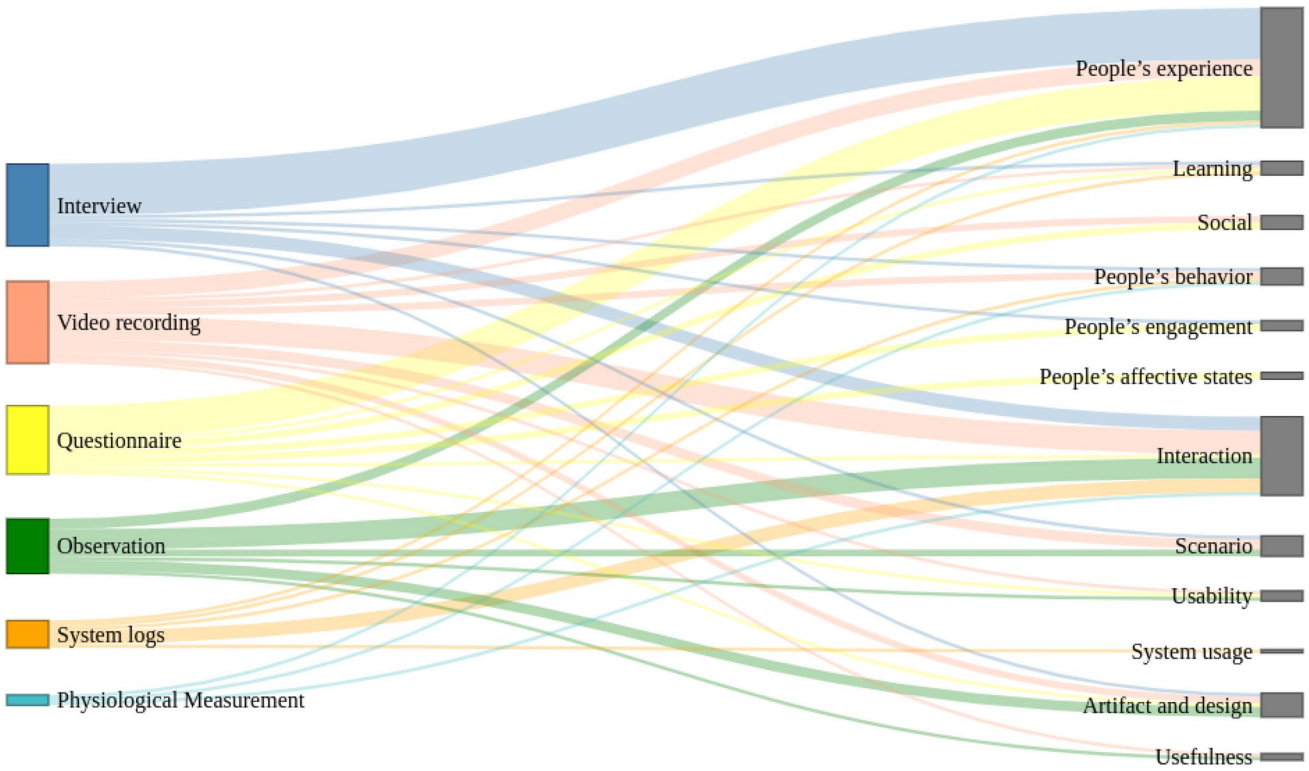

Fig. 6 Relations between data collection method categories (left) and objects of evaluation (right)

usefulness of systems, and social aspects, whereas the interview was also used to assess the engagement of people with technology. Similarly, *questionnaires* were applied to capture information about the experience, engagement, learning and social relations of people, the usability of systems, the people's interaction, and the evaluation of artifacts and their design. Unlike the interviews and video recordings, the questionnaires were also used to collect information about the affective states of people. *Observations* (note taking) were mainly used to collect information about the people's experience and their interactions and to assess the artifacts and their design. *System logs* and *physiological measurements* were mainly used to capture data from the experience and behavior of people, and their interactions. Overall, interviews and questionnaires were used mainly to collect data about the people's experience, whereas video recordings and observations were used mainly to capture interactions of people with the environment and with others.

One aspect that is illustrated in Fig. 6 is the asymmetry between the use of subjective and objective data collection methods categories, with a substantial preference for subjective ones. Real-time observations, or a posteriori through video recordings, as well as interviews and questionnaires, are data collection methods with an intrinsic subjectivity in them: the researchers have an active role in proposing them, and/or the participants are reminiscing about their experience and trying to translate it into a verbal or some other proposed format. More objective approaches, such as system logs and physiological measurements, although present, represent only a small fraction of the data collection method categories found in our results. We wonder whether it is not due to the lack of evaluation methods informed by objective data. The other side of Fig. 6, in turn, illustrates some balance between more general and more specific objects of evaluation. More general objects of evaluation (e.g., people's experience and interaction) are represented in a little over half of the selected studies, while the other more specific subjects of evaluation (e.g., learning, social, usability), although fragmented in several categories, compose what amounts to almost half of the selected studies.

It is important to emphasize here that our argument is not that objective data collection methods should be favored over subjective ones, nor is one better than the other, as both can be conducted with adequate scientific rigor and yield important results. Neither should general or more specific objects of evaluation be mutually exclusive, as both can play an integral role in the research objectives of a study. We perceive, however, that a balance between subjective and objective data collection methods could benefit evaluation in scenarios of ubiquity of technology, as complementary methods have the potential of broadening both the depth and range of what is being evaluated.

For RQ3, "Are evaluation methods taking into account the new types of (social) actions and perceptions afforded by interactive installations? If so, how?" Having observed the *what* of evaluation in RQ2, we can finally examine *how* those aspects are evaluated by the selected studies. Looking at the main categories of instruments reported in the studies and described in Sect. 3.3, we observe that interviews and questionnaires are predominant, totaling 46,7%, followed by video recordings with 21.3% of presence. In particular, one interview method (video-cued recalls [60]) stands out to evaluate embodied experiences by employing video recordings as a support to guide the interview. Moreover, an interview method (second-person interview [63]) to get detailed accounts from participants as close as possible to their first-person experiences was also present in our results.

Considering the enactive approach to cognition, it emphasizes the extended, intersubjective, and socially situated nature of cognitive systems [51]. According to Gallagher et al. [51], "the world (meaning, intentionality) is not pregiven or predefined but is structured by cognition and action". When considering the relevance of social action and perception in ubiquitous environments, we observe that the works have addressed aspects such as interaction, awareness, coordination, and collaboration. Although some works recognize the relevance of social aspects in evaluation [21, 32, 36, 45], specific methods for evaluating the environments regarding their affordances for social action and perception, embodiment, social sense-making, etc. are hardly found among the works. In [45], for example, visitors' social interactions in a public setting have been investigated through video observations of interactions with a musical tabletop. It is well known that tabletop interfaces support social interaction and learning. In [21], the authors recognized the importance of embodiment in social interaction and social cognition, and they proposed a questionnaire to evaluate the perceived collaboration and how group members feel about each other. Similarly, in [36], the authors applied a questionnaire to know how participants felt socially and how aware they were of others. Although these works evaluate social interaction, none of them considers the intersubjective relations (e.g., joint action, joint attention, and participatory sense-making) that emerge in the ongoing interactions on ubiquitous scenarios. Furthermore, in [32], we found a framework with categorical dimensions like "visitor experience" and "social setting" to analyze interactive media arts. The "social setting" category contains several attributes to examine and evaluate whether an installation scales to multiple visitors, requires coordination, etc. As a result of the analysis of these works, we identified relevant aspects to be considered by evaluation methods like awareness of the others, social affective understanding, emotional interdependence, and intersubjective relations. However, it is not enough to just consider these aspects in the evaluation methods, but it is necessary to address them in the design of ubiquitous scenarios to promote situations of social, affective, and intersubjective relations.

4.2 Highlights for a research agenda

Informed by the results presented and discussed so far in this article, overall, important aspects of interaction of people in these environments are somehow being evaluated; e.g., users' experience, engagement, behavior, affective states. Nevertheless, missing aspects are also visible and raise opportunities for new research agendas, briefly summarized as follows.

Experience as a concept.

Although 37% of the studies in our final selection reported some kind of evaluation of people's experiences, it is noticeable how their notion of experience differs from HCI's mainstream concept of user experience (UX). While UX can be considered a jargon with a disputed and not so well-defined meaning within the field, the notion of experience found in our results instead seems to be based on a phenomenological perspective. Studies [20, 28, 39], for instance, draw on Merleau-Ponty's seminal book "Phenomenology of Perception" to establish the importance of the role a person's body in the experience of interacting with the world and making sense of it, therefore characterizing a notion of experience that is intrinsically embodied. This amounts to a notion of experience that is not based entirely on subjectivity and ill-defined feelings, as the case of UX, but instead is grounded on the fact that whatever we may experience, we always do it through our bodies and its distinct sensorimotor capabilities. This embodied notion of experience is also evidenced in selected studies with a background that also originates from phenomenology, such as the theories labeled as 4E (embodied, embedded, enacted, or extended) cognition (e.g., [12, 29]).

In sum, although people's experience appears with relevance in the works, the research community could benefit from going further in the types of experience afforded by those environments. In the same way, while affective states of people interacting in the studied scenarios are already being presented by some works, this phenomenon could be further located within a phenomenological frame to consider some important aspects of affectivity related to their bodies, their action, their embodied actions, and in their experience with artifacts and with others.

Multiplicity and reachness of methods.

As presented in Sect. 3.3, studies of the SRL use a variety of data collection methods. The analysis of the 45 selected studies shows that 20 studies (44.44%) made use of more than the one data collection method, with some of them making use of as many as four methods (e.g., [8, 25, 38]). We interpret

this number as a hint that a significant number of researchers do not consider a single-data collection method to be enough to address evaluation in scenarios of ubiquitous and pervasive technology. For instance, we found studies using video recordings and system logs to obtain information about people's interactions (e.g., [3, 8, 25]) or interviews complemented with video recordings to evaluate the people's experience (e.g., [10, 15, 38]). Ubiquitous and pervasive scenarios have a wide range of dimensions and aspects to be considered in an evaluation process, so it follows that it is unlikely for a single known method to provide sufficient insights regarding a scenario in its entirety and complexity. Nevertheless, frameworks which structure and organize the methods into different dimensions are hardly found and would be helpful.

While different methods to collect data may be useful as an important resource for researchers to explore multiple aspects of ubiquitous and pervasive systems in a complementary way, they still may not be enough to address an ever growing list of possible dimensions and aspects to be considered in the design and evaluation of ubiquitous and pervasive systems. For example, social aspects of the interaction, in the sense of its intersubjective aspects, e.g., the joint attention of people through interaction, their joint coordination, and the way they could enrich a participatory sense-making, are all aspects of design and evaluation of ubiquitous and pervasive scenarios that are open as new research agendas.

Ethical issues It is noticeable that despite evaluation in HCI relies heavily on people's participation, only 11 out of the 45 works (24.44%) explicitly mention either a research ethics committee approval for their study, or at least informed consent from the participants. Although we cannot assess individually whether this information is just not made explicit in the full texts or there was no research ethics committee and consent forms involved at all, we consider this a low number, suggesting that this ethical aspect might have been overlooked by researchers, peer-reviewers, and publication vehicles, in the set of studies analyzed. Moreover, as the main types of interaction in the ubiquitous environments address mainly embodied, full-body, motion, and gesture-based types of interaction with technology, more concern is raised regarding not only the people's participation in evaluation methods, but also aspects of design of those environments, which impact on safety and privacy issues, accessibility, and personal values.

5 Conclusion

From computer use limited to performing well-defined tasks, we are now experiencing computational technologies pervading many areas of our lives, in all places, and being used simultaneously by different people. Ubiquitous and pervasive technologies have created new forms of interaction, demanding new types of approaches for the design and evaluation of systems. This work has addressed this context by inquiring about the evaluation methods being practiced. A systematic literature review on scenarios of ubiquitous and pervasive systems, focusing on the context of interactive installations was conducted to shed light on the subject.

Results of the systematic literature review revealed that people's experience, understood in its wider sense, was the most highlighted subject of evaluation. Moreover, the concept of "experience" and its different interpretations, carried by the works, still demands further investigation. Interviews were the most practiced category of data capture method, revealing predominantly self-reported data which summed up to 46.7% of the works, when added to the questionnaire category. Our study has also exposed a combination of methods being applied in the studies, suggesting the complexity of the phenomena involved in the interaction in these environments and the inability of a single method to cover aspects of this complexity. While our results refer to the context of interactive installations in exhibition spaces, other areas of ubiquitous and pervasive technology environments, such as for example IoT (Internet of Things) scenarios, ambient intelligence, social wearable artifacts, to name a few, still could add specificities to the results already pointed out. Further work involves addressing the open research agenda, especially considering the overarching concepts of phenomenology for addressing experience, emotion, values, and the reliability of methods.

Funding This study was financially supported by the National Council for Scientific and Technological Development (CNPq) through grant #306272/2017–2, by the São Paulo Research Foundation (FAPESP) through grants #2015/16528–0, #2015/24300–9, #2020/03503–7, and #2020/04242–2, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflicts of interest Not applicable.

Systematic literature review references

1. Adhitya S, Scott D (2018) The London Soundmap: integrating sonic interaction design in the urban realm. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion* (New

York, NY, USA, 2018), AM'18, Association for Computing Machinery.
2. Afonso AG, Ergin E, Fatah G, Schieck A. (2019) Flowing bodies: exploring the micro and macro scales of bodily interactions with urban media installations. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (New York, NY, USA, 2019), DIS'19, Association for Computing Machinery, 1183–1193.
3. Batras D, Guez J, Jégo J-F, and Tramus M-H. (2016) A virtual reality agent-based platform for improvisation between real and virtual actors using gestures. In *Proceedings of the 2016 Virtual Reality International Conference* (New York, NY, USA, 2016), VRIC '16, Association for Computing Machinery.
4. Biloria N, Dritsa D (2020) Real-time interactive multimodal systems for physiological and emotional wellbeing. Springer International Publishing, Cham, pp 181–203
5. Brown C, Paine G (2019) A case study in collaborative learning via participatory music interactive systems: Interactive Tango Milonga. Springer International Publishing, Cham, pp 285–306
6. Ceconello M, and Spallazzo D (2016) Vis-à-vis with Leonardo. Designing Digital Encoun-ters. In *Eurographics Workshop on Graphics and Cultural Heritage* (2016), C. E. Catalanoand L. D. Luca, Eds., The Eurographics Association.
7. Cuan C, Berl E, LaViers A (2019) Time to compile: a performance installationas human-robot interaction study examining self-evaluation and perceived control. Paladyn, J Behav Robotics 10(1):267–285
8. Dalsgaard P, Dindler C, Halskov K (2011) Understanding the dynamics of engaging interaction in public spaces. In: Campos P, Graham N, Jorge J, Nunes N, Palanque P, Winckler M (eds) *Human-Computer Interaction - INTERACT 2011* (Berlin, Heidelberg, 2011). Springer, Berlin, Heidelberg, pp 212–229
9. Erkut C, Fehr J (2017) Structuring design and evaluation of an interactive installation through swarms of light rays with human-artifact model. In: Brooks AL, Brooks E (eds) *Interactivity, Game Creation, Design, Learning, and Innovation* (Cham, 2017). Springer International Publishing, pp 39–46
10. Françoise J, Candau Y, Fdili Alaoui S, and Schiphorst T. (2017) Designing for kinesthetic awareness: revealing user experiences through second-person inquiry. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA,2017), CHI '17, Association for Computing Machinery, 5171–5183.
11. Froese T, Suzuki K, Ogai Y, Ikegami T (2012) Using human–computer interfaces to investigate 'mind-as-it-could-be' from the first-person perspective. Cogn Comput 4(3):365–382
12. Grace K, Grace S, Maher ML, Mahzoon MJ, Lee L, LoCurto L, and Outcault B. (2017) The willful marionette: exploring responses to embodied interaction. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition* (New York, NY, USA, 2017), C&C '17, Association for Computing Machinery, 15–27.
13. Hsueh S, Alaoui SF, and Mackay WE (2019) Understanding kinaesthetic creativity in dance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, 1–12.
14. Ingebritsen R, Knowlton C, Sato H, Mott E. (2020) Social movements: A case study in dramaturgically-driven sound design for contemporary dance performance to mediate human-human interaction. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY, USA, 2020), TEI '20, Association for Computing Machinery, 227–237.
15. Janauskait L, and Palamas G. (2019) Establishing dialogues between movement and atmospheric ambiances. In *Proceedings of the 6th International Conference on Movement and Computing* (New York, NY, USA, 2019), MOCO '19, Association for Computing Machinery.
16. Kelton ML, Ma JY (2018) Reconfiguring mathematical settings and activity through multi-party, whole-body collaboration. Educ Stud Math 98(2):177–196
17. Levinsky S, and Russell A. (2019) Agency in dialogue how choreographic thought emerges through dancing with tools that propel. In *2019 AISB Convention (2019), The Society for the Study of Artificial Intelligence and Simulation of Behaviour*, 7–14.
18. Liang R-H, Chung W-M, Kao H-L, and Lin T-Y (2013) Intouch: crossing social interaction with perception. In: Marcus A (ed) Design, User Experience, and Usability. User Experience in Novel Technological Environments (Berlin, Heidelberg, 2013). Springer, Berlin, Heidelberg, pp 306–315
19. Loke L, Khut G P, Kocaballi AB. (2012) Bodily experience and imagination: designing ritual interactions for participatory live-art contexts. In *Proceedings of the Designing Interactive Systems Conference* (New York, NY, USA, 2012), DIS '12, Association for Computing Machinery, 779–788.
20. Lopes MM (2016) Body storytelling and the performance of memory: arts-based-research and human enhancement. In *Distributed, Ambient and Pervasive Interactions* (Cham, 2016), N. Streitz and P. Markopoulos, Eds., Springer International Publishing, 257–269.
21. Malinverni L, Burguès NP. (2015) The medium matters: the impact of full-body interaction on the socio-affective aspects of collaboration. In *Proceedings of the 14th International Conference on Interaction Design and Children* (New York, NY, USA, 2015), IDC '15, Association for Computing Machinery, 89–98.
22. Manzolli J, Moroni A, Valarini GA. (2018) Selfhood: an evolutionary and interactive experience synthesizing images and sounds. In *Music Technology with Swing* (Cham,2018), M. Aramaki, M. E. P. Davies, R. Kronland-Martinet, and S. Ystad, Eds., Springer International Publishing. 625–636.
23. Mendoza YLM, Baranauskas MCC. (2019) Tangitime: designing a (socio)enactive experience for deep time in an educational exhibit. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems* (New York, NY, USA, 2019), IHC '19, Association for Computing Machinery.
24. Millar GC, Tabrizian P, Petrasova A, Petras V, Harmon B, Mitasova H, Meetenmeyer RK (2018) Tangible landscape: a hands-on method for teaching terrain analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, 1–12.
25. Mora-Guiard J, Crowell C, Pares N (2018) A play therapy based full-body interaction intervention tool for children with autism. In *Pervasive Computing Paradigms for Mental Health* (Cham, 2018), N. Oliver, S. Serino, A. Matic, P. Cipresso, N. Filipovic, and L. Gavrilovska, Eds., Springer International Publishing, 28–40.
26. Mora-Guiard J, Pares N. (2014) "Child as the measure of all things": the body as a referent in designing a museum exhibit to understand the nanoscale. In *Proceedings of the 2014 Conference on Interaction Design and Children* (New York, NY, USA, 2014), IDC '14,Association for Computing Machinery, 27–36.
27. Morgan E, Gunes H (2013) Human nonverbal behaviour understanding in the wild for new media art. In: Salah AA, Hung H, Aran O, Gunes H (eds) *Human Behavior Understanding* (Cham, 2013). Springer International Publishing, pp 27–39
28. Núñez-Pacheco C, Loke L (2014) Aesthetic resources for technology-mediated bodily self-reflection: the case of eloquent robes. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (New York, NY, USA, 2014), OzCHI '14, Association for Computing Machinery, 1–10
29. Prpa M, Tatar K, Riecke BE, Pasquier P (2017) The pulse breath water system: exploring breathing as an embodied interaction for enhancing the affective potential of virtual reality. In: Lackey S, Chen J (eds) *Virtual, Augmented and Mixed Reality* (Cham, 2017). Springer International Publishing, pp 153–172

30. Roberts J, Lyons L (2017) The value of learning talk: applying a novel dialogue scoring method to inform interaction design in an open-ended, embodied museum exhibit. Int J Comput-Support Collab Learn 12(4):343–376
31. Salmon R, Paine G (2013) Embodiment: auditory visual enhancement of interactive environments. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (New York, NY, USA, 2013), TEI '13, Association for Computing Machinery, 129–136.
32. Schacher JC, Bisig D (2017) Haunting space, social interaction in a large-scale media environment. In: Bern-haupt R, Dalvi G, Joshi A, Balkrishan DK, O'Neill J, Winckler M (eds) *Human-Computer Interaction - INTERACT 2017* (Cham, 2017). Springer International Publishing, pp 242–262
33. Schiphorst T. (2011) Self-evidence: applying somatic connoisseurship to experience design. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI EA '11, Association for Computing Machinery, 145–160.
34. Shaer O, Valdes C, Liu S, Lu K, Chang K, Xu W, Haddock TL, Bhatia S, Densmore D, Kincaid R (2014) Designing reality-based interfaces for experiential bio-design. Pers Ubiquit Comput 18(6):1515–1532
35. Simon LU, van der Vlugt M, Calvi L. (2016) Triggers to entice an audience to 'perform as interface' in an interactive installation. In *Proceedings of the 20th International Academic Mindtrek Conference* (New York, NY, USA, 2016), Academic Mindtrek '16, Association for Computing Machinery, 322–330.
36. Stepanova ER, Desnoyers-Stewart J, Pasquier P, Riecke BE (2020) Jel: Breathing together to connect with others and nature. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (New York, NY, USA, 2020), DIS '20, Association for Computing Machinery, 641–654.
37. Tan L, and Chow KKN (2017) Facilitating meaningful experience with ambient media: an embodied engagement model. In *Proceedings of the Fifth International Symposium of Chinese CHI* (New York, NY, USA, 2017), Chinese CHI 2017, Association for Computing Machinery, 36–46.
38. Tanenbaum K, Hatala M, Tanenbaum J, Wakkary R, Antle AA (2014) case study of intended versus actual experience of adaptivity in a tangible storytelling system. User Modeling User-Adapted Interact 24(3):175–217
39. Tikka H, Na SV, Jacucci G, Korpilahti T (2011) Provoking the city–touch installations for urban space. Digital Creativity 22(3):200–214
40. Torpus J. (2018) Extending museum exhibits by embedded media content for an embodied interaction experience. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (New York, NY, USA, 2018), NordiCHI '18, Association for Computing Machinery, 236–246.
41. Vets T, Nijs L, Lesaffre M, Moens B, Bressan F, Colpaert P, Lambert P, Van de Walle R, Leman M (2017) Gamified music improvisation with billiart: a multimodal installation with balls. J Multimodal User Interfaces 11(1):25–38
42. Vidyarthi J, Riecke BE, and Gromala D. (2012) *Sonic Cradle*: designing for an immersive experience of meditation by connecting respiration to music. In *Proceedings of the Designing Interactive Systems Conference* (New York, NY, USA, 2012), DIS '12, Association for Computing Machinery, 408–417.
43. Volpe G, Camurri A (2011) A system for embodied social active listening to sound and music content. J Comput Cult Herit 4:1
44. Wiethoff A, Gerstberger J, Gehring S. (2015) Starlight: exploring embodied interactions with media architecture and public audiences. In *Proceedings of the 4th International Symposium on Pervasive Displays* (New York, NY, USA, 2015), PerDis '15, Association for Computing Machinery, 83–89.
45. Xambó A, Hornecker E, Marshall P, Jordà S, Dobbyn C, Laney R (2016) Exploring social interaction with a tangible music interface. Interacting with Comp 29(2):248–270

References

46. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 25(1):49–59. https://doi.org/10.1016/0005-7916(94)90063-92
47. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3(2):77–101. https://doi.org/10.1191/1478088706qp063oa
48. Brockmyer JH, Fox CM, Curtiss KA, McBroom E, Burkhart KM, Pidruzny JN (2009) The development of the game engagement questionnaire: A measure of engagement in video game-playing. J Exp Soc Psychol 45(4):624–634. https://doi.org/10.1016/j.jesp.2009.02.016
49. Costello B, Edmonds E. (2007) A study in play, pleasure and interaction design. In: Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces. 76–91. DPPI '07, ACM, New York, NY, USA. https://doi.org/10.1145/1314161.13141685.
50. Díaz-Oreiro I, López G, Quesada L, Guerrero LA (2019) Standardized questionnaires for user experience evaluation: a systematic literature review. Proceedings 31(1). https://doi.org/10.3390/proceedings2019031014
51. Gallagher S, Lindgren R (2015) Enactive metaphors: learning through full-body engagement. Educ Psychol Rev 27:391–404. https://doi.org/10.1007/s10648-015-9327-1
52. Gough D, Oliver S, Thomas J. (2012) An introduction to systematic reviews. SAGE Publications
53. Harms C, Biocca F. (2004) Internal consistency and reliability of the networked minds measure of social presence. In: Seventh Annual International Workshop: Presence, http://cogprints.org/7026/
54. Hassenzahl M, Burmester M, Koller F (2003) Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In: Szwillus G, Ziegler J (eds) Mensch & Computer 2003: Interaktion in Bewegung. Vieweg Teubner Verlag, Wiesbaden, pp 187–196. https://doi.org/10.1007/978-3-322-80058-919
55. Höök K, Sengers P, Andersson G (2003). Sense and sensibility: evaluation and interactive art. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). ACM, 241–248. https://doi.org/10.1145/642611.642654
56. Kaipainen M, Ravaja N, Tikka P, Vuori R, Pugliese R, Rapino M, Takala T (2011) Enactive systems and enactive media: embodied human-machine coupling beyond interfaces. Leonardo 44(5):433–438. https://doi.org/10.1162/LEONa00244
57. Korhonen H, Montola M, Arrasvuori J (2009) Understanding playful user experience through digital games. Int Conference Designing Pleasur Prod Interfaces 09:274–285
58. Lazar J, Feng JH, Hochheiser H (2010) Research methods in human-computer interaction. Wiley Publishing
59. Lucero A, Arrasvuori J. (2010) Plex cards: a source of inspiration when designing for playfulness. In: Proceedings of the 3rd International Conference on Fun and Games. 28–37. Fun and Games '10, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/1823818.1823821
60. McLennan J, Omodei M, Wearing A. (2001) Cognitive processes of first-on-scene fire officers in command at emergency incidents as an analogue of small-unit command in peace support operations. In: The human in command peace support operations, 312–329
61. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Medicine 6(7):1–6. https://doi.org/10.1371/journal.pmed.1000097
62. O'Brien HL, Toms EG (2008) What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inform Sci Technol 59(6):938–955. https://doi.org/10.1002/asi.20801

63. Petitmengin C (2006) Describing one's subjective experience in the second person: an Interview method for the science of consciousness. Phenomenol Cogn Sci 5(3):229–269. https://doi.org/10.1007/s11097-006-9022-2
64. Smith J, Flowers P, Larkin M (2009) Interpretative phenomenological analysis: theory, method and research. SAGE Publications
65. Torpus J (2018) Extending museum exhibits by embedded media content for an embodied interaction experience. In Proceedings of the 10th Nordic Conference on Human-Computer Interaction (New York, NY, USA, 2018), NordiCHI '18, Association for Computing Machinery, 236–246.
66. Varela F, Thompson E, Rosch E (1993) The embodied mind: cognitive science and human experience. Philosophy, psychology, MIT Press, Cognitive science
67. Vermersch P (1996) L'explicitation de l'action. Cahiers de linguistique sociale 3:113–120
68. Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 54(6):1063–1070
69. Weiser M (1991) The computer for the 21st century. Sci Am 265(3):94–104