Remote Usage Data Collection and Analysis for Mobile Accessibility Applications

Dragan Ahmetovic, Cristian Bernareggi, Mattia Ducci, Andrea Gerino, Sergio Mascetti
Department of Computer Science, University of Milan, Italy
Email: [name].[surname]@unimi.it

***Abstract*—Conducting experimental research in the field of mobile accessibility and assistive technologies is difficult due to the low numerosity of the representative population. To address this issue, a possible approach is remote collection and analysis of usage data through publicly available mobile applications. This method is useful for performing large scale evaluations and acquiring knowledge of the target population and their behavior. The acquired knowledge can be used to advance future research and to improve the mobile applications themselves.**

***Index Terms*—Remote logging, mobile assistive technologies.**

I. Introduction

Research in the field of assistive technologies requires to conduct experiments involving people with disabilities [1]. A common approach is to conduct supervised user studies, with the participants located in the same place as the supervisor (in-situ). This approach has a number of advantages. In particular, the researchers can directly observe the participants, the experimental scenario is controlled, and it is possible to immediately intervene in case support is needed during the experiments (*e.g.*, the participant is unable to use the system being tested). Despite these advantages, supervised user studies also have limitations. They imply a geographic bias since all participants come from the same area. Candidate participants with disabilities are often few due to strict inclusion criteria (*e.g.*, only blind people) and difficulties in reaching the experimental site [1,2]. Involving participants for more than one session is also difficult, so longitudinal studies are uncommon. Direct observation may also influence study results due to “Hawthorne Effect” [3]. Finally, extraordinary circumstances, such as the COVID-19 pandemic, can limit the possibility to interact with participants in person [4].

We share our experience with a different user evaluation approach in the field of assistive technologies. Our approach is based on remote collection of usage data through applications that are made available to the end users, either publicly (*e.g.*, from mobile app stores), or for testing purposes (*e.g.*, using TestFlight on iOS devices [5]). This way, it is possible to persistently and pervasively collect traces of usage behavior at a large scale, without geographical constraints. The collected data can then be used to evaluate the application, and also explore usage behavior and user interactions with it, possibly yielding more general results than experiments conducted with a small number of local participants. We discuss the advantages and disadvantages of this approach, and we briefly introduce the technical tools that we used to implement it.

II. Related Work

We review prior work discussing key limitations of supervised, in-situ user studies in the field of assistive technologies. Then, we survey related work on studies using remote usage data, in particular addressing mobile device applications.

A. User Studies in Assistive Technologies Research

Assistive technologies research is a multidisciplinary effort. Considering the specific problems addressed, the methodologies involved may intersect different research fields, including pervasive and mobile computing, data management, distributed systems, and others. In most cases, however, evaluating assistive technologies involves Human-Computer Interaction research methods such as surveys [6] and user studies [7]. In particular, supervised user studies are invaluable to capture real experiences of end-users [8]. Such studies should be conducted with representative participants [1] to ensure their validity.

However, it is challenging to recruit many participants for a user study [7], and participants with disabilities in particular are often few [9]. Indeed, studies frequently have strict inclusion criteria, limiting possible participants to a small (sub)population [1]. In such cases only a few candidates might be available locally, and those with sensory, motor or cognitive impairments might also have difficulties to travel to the experimental site, further reducing participants number [2].

A general limitation of supervised user studies is that, since they can be conducted only with local participants, they introduce a geographical bias, potentially limiting the generalizability of the results [10]. Indeed, cultural and policy factors were shown to influence the results in assistive technology research [11]. Longitudinal studies are also rare [12], because time and effort requirements of such studies are a barrier in recruiting potential participants. In supervised experiments, participants’ behavior can also be influenced by the knowledge that they are in a study (“Hawthorne effect”) [3], or that they are evaluating a system made by the experimenters, leading to positive Response Bias [13].

B. Studies with Remote Usage Data

Studies involving remote usage data collection overcome the above issues because they can be conducted at large scale, without geographical limitations or the requirement to travel to a specific experimental site [2]. In particular, this approach has been used for remote evaluation of mobile applications [14], enabling longitudinal studies on the usage of mobile services and applications [15]. The collected usage data can be analyzed to detect common interaction patterns and inform interface personalization [16]. It has also been used for large-scale clustering of mobile applications by category [17]. Such analysis can also be conducted together with the users who can provide additional information about the recorded activity [18]. General purpose frameworks have also been proposed to support usage data recording and collection of subjective responses from the users [19]. In this work we consider the problem from the specific point of view of remote usage data studies with assistive technologies.

III. Remote Study Design Methodology

We describe the study design methodology adopted for remote experiments. In particular, we outline three approaches for conducting remote studies, and we discuss the key challenges that need to be addressed as a part of this study design.

In the following we consider two actors: the researcher, who wants to test a software system or an *application*, and the participant, who is typically a person with a disability or a stakeholder (*e.g.*, a teacher, a caregiver, etc...). The methodology is based on the idea that the participant uses the application, which collects usage data and stores it on a remote database. Usage data is associated to a pseudo-identifier *i.e.,* a unique code assigned to each participant, and contains information about the device used, relevant system settings (*e.g.*, accessibility features used), and application interactions performed by the users. The collected usage data is afterwards analyzed by the researcher.

A. Experimental Approaches

We present three key experimental approaches for implementing remote studies which we explored in our past work.

Approach 1: Implicit Experiments [20]–[23]: the app to be evaluated is made publicly available (*e.g.*, through online app stores or as web-app). The participant uses the application to benefit from its key functionalities, for example because the application helps them solving an everyday problem or it entertains them, like in the case of games. The participant is aware that interaction data will be collected, because privacy-related information is provided within the user agreement, but they do not perceive app usage as an experiment.

Approach 2: Explicit Unsupervised Experiments [24]–[27]: the application to be evaluated is made available to a set of participants, which may either be known or recruited anonymously (*e.g.*, through snowball sampling). The participants are asked to run the app, typically with a target task (*e.g.*, follow instructions until a given message is shown, or play at least four levels of a game). In this case the participants are aware that they are running an experiment. The motivation is provided by the researcher, for example in the form of monetary compensation or as an implicit incentive for social good. The expected duration of the experiment by each participant can be predicted given the expected task. The actual duration of each test can be verified ex-post (*i.e.*, when analyzing the results).

Approach 3: Explicit Supervised Experiments [28]: this approach is similar to the ***Explicit Unsupervised Experiment***, with the main difference that the researcher supervises the experiment remotely, for example on the phone or through a video call. This approach requires an additional effort from the researcher, in terms of the time required to setup the remote meeting, and the time spent for the experiment itself. Also, there might be additional technical problems, for example due to non-accessible video conferencing tools. One consideration is that the participant should run the (video) call from a different device than the one used for the experiment because there are situations in which the app being tested can have a different behaviour during a call (*e.g.*, for the reproduction of audio). This limitation, however, requires the participant to have another device and to be able to set up a call with it. The advantage of this technique is that the researcher can observe, or at least listen to the participant, and this often discloses a number of insights. Also, being able to directly interact with the participant, the researcher can solve problems or provide feedback (*e.g.*, in case the participant is stuck and cannot complete the task). On the other hand, as for in-situ experiments, this supervision may influence the results [3].

B. Challenges

Remote study design approaches entail a number of challenges. We discuss the four major ones.

*1) **Recruitment**:* in Approach 3, the researchers actively recruit the participants, which is a time consuming task. Instead, in Approach 2, participants can either be recruited directly, or the study can be advertised on thematic forums or through mailing list. Finally, in Approach 1, the researchers do not directly recruit the participants, but they still might facilitate the diffusion of the app by publicising it, as in the previous case. In all cases, the collaboration with associations and NPOs (Non-Profit Organizations) can facilitate the process by stimulating the interest of end users or stakeholders. For example, our *Musa* application [1] was tested with Approach 2, and one association[2] helped us recruiting the participants. Similarly, another association[3] provided support in advertising our application *iMove around* [20].

*2) **Participants' selection**:* this is another challenge strictly related to the recruitment. In Approaches 2 and 3 the researcher can define some **inclusion criteria** and apply them during the selection for example to involve legally blind participants only. Instead, in Approach 1 it is not possible to predict who will use the application. To overcome this limitation, it is possible to collect data so that, during the analysis phase, it is possible to profile users, hence identifying the target population ex-post. This is for example what we did with our research with *iMove around*, in which we identified users with visual impairments by collecting system settings about visual enhancements and screen reader usage [21].

[1]https://www.descrivedendo.it/musa/il-progetto/

[2]ANS - Associazione Nazionale Subvedenti (National Association of People with Low Vision).

[3]Retina Italia Onlus, an association of people with retinitis pigmentosa.

*3) **Motivating participants**:* is another challenge that differs significantly in the three Approaches. Indeed, in Approach 1 the participants use the application because it is useful or entertaining for them. So, beyond recruiting the participants (*i.e.*, bringing them to install and run the application a first time) it is also important to motivate them to continue using the application, so that sufficient information can be collected. We experienced the challenge of this with our *Invisible Puzzle* application, which is a game aimed at testing sonification techniques for touchscreen interaction [27]: analyzing the data collected with Approach 1, we realized that many users stopped using the application after the first levels, hence providing only partial data for analysis. The challenge is different in Approaches 2 and 3, in which the participants are motivated by the fact that they agreed to participate to a test, hence they are more inclined to complete it. Still, since there is no supervision, we observed that some participants do not conclude the test, and we suspect this might due to the fact that are unable to do it, or they find it too mentally demanding or time consuming.

*4) **Application Engineering**:* for experiments conducted in-situ, a common practice is to use application prototypes that only need to run on the device used for the experiment. Instead, remote data collection requires more engineered applications because they need to be easy to install and use by the participants on their own devices. If this does not hold, there is an implicit selection criteria: participants that do not have access to a given set of devices are excluded from the test. In particular, for Approach 1, applications need to be of sufficient quality to be approved on app distribution platforms, and they should also run on the major mobile platforms (iOS and Android), possibly also on less recent versions of the operating system.

C. Implementation aspects

There are two main aspects to consider when implementing the remote data collection methodology presented in this paper.

*1) **Which Data to Collect**:* we want to collect all relevant user interactions and other information that can be used to derive participant characteristics, like the fact that participant is interacting with the application through a screen reader. Information is organized into log records, each one representing an interaction between the participant and the app. For each log record it is also necessary to store its timestamp, device information (*e.g.*, hardware, operating system), system configuration (*e.g.*, language, accessibility services), application data (name and version), and the participant pseudo-identifier. This last information is particularly challenging to collect; technically, it is possible to obtain a code that identifies the pair $\langle$application, device$\rangle$. While this code conceptually represents a participant, there are two cases that are hard to trace. First, if a user uninstalls and reinstalls the app, a new code is generated, so the user will result as a new one. Similarly, if a person uses two or more devices, the codes are different and are hard to link to the same person. A possible workaround for these problems is to have an application that requires the user to explicitly register and log in. Another problem occurs when the same device is shared by more people. For example, in *MathMelodies* [23] and *WordMelodies* [22] this happens when a caregiver and a child alternate in using the application (*e.g.*, the caregiver explains to the child how to use the application or how to solve an exercise). In these cases it is a challenge to distinguish which user is actually using the application.

When designing which data to collect, it is also important to take into account **ethical and privacy issues**. In particular, in our research we do not collect information that can be used to re-identify the user. This means, for example, that we do not collect location information (or any location-related information), images from the camera or recordings from the microphone. Also, it is important to inform the user about which data will be collected.

*2) **Data Collection System**:* while third party logging systems exist, due to privacy concerns involving recording and storage of sensitive data, we developed an in-house tool, called *Icarus*, that runs and stores logs on a server managed by the researchers. As shown in Figure 1, *Icarus* is composed of two main components: on the server side a RESTful [29] web service (written in *NodeJS*[4] and *Express*[5]) receives the logs and stores them on a no-sql database (*MongoDB*[6]). We chose to adopt a no-sql database to support logging and querying data without a fixed schema, thus enabling us to customize the recorded logs for each application that uses the service. On the client side, various SDKs (Software Development Kits) have been developed for different programming languages and platforms (native iOS and Android, React Native, Flutter, etc...). Each SDK can be easily integrated with the applications and exposes a main function that the app can invoke to log data. The function performs three main tasks: first, it enriches each log request with the required metadata (i.e., timestamp, application name and version, etc.); second, it locally stores log records, in case a network connection is not available; third, it transmits the logs to the server, when a connection is available. *Icarus* source code, together with installation instructions, is available online[7].

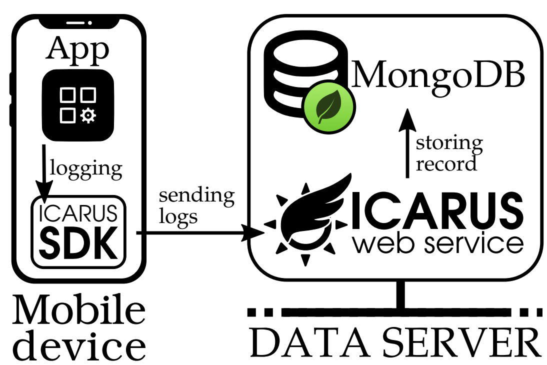

Fig. 1: Icarus system architecture

[4]https://nodejs.org
[5]https://expressjs.com
[6]https://www.mongodb.com
[7]https://ewserver.di.unimi.it/gitlab/icarus

IV. Data Analysis Methodology

Once collected, the remote usage data is analyzed to evaluate the app and to comprehend participants' behavior while using it. We define three types of analyses of the considered data based on the temporal dimension of the analysis.

A. Demographic and Experimental Setting Analysis

Unlike supervised studies in a controlled scenario, which are commonly conducted on a same device, remote studies are conducted on participants' own devices, with personalized settings and preference configurations. On the one hand, the variety of different experimental settings may affect the study results and therefore needs to be considered as a factor. On the other hand, analyzing participants' demographics and the experimental setting allows us to understand real-world conditions of mobile device usage. Since demographic and device data do not change, this type of analysis is conducted without considering the temporal dimension of the logs. However, user settings might change (*e.g.,* screen reader activation), and therefore may need to be considered during *Usage Behavior Analysis*. In particular, we assess three key factors.

1) ***User Abilities and Disability****:* by collecting information on assistive services active on the mobile device, we can infer the range of participants' abilities and the characteristics of their disability. Indeed, differences in individual abilities can be reflected in system settings (*e.g.*, screen reader usage and speed, magnification factor). This may help to better frame the app target users, sometimes unveiling unintended user segments. For example, *iMove around* was intended for users who are blind. Instead, almost 75% of users did not always use a screen reader, indicating that users with residual vision also find the system useful [20].

2) ***Interaction Preferences and Personalization****:* users may display different interaction preferences with the system across different user segments, and with respect to the app defaults. Analyzing the changes in user preferences and their distribution can help the researchers to revisit system defaults or to pinpoint those settings that may be often changed and therefore need to be easy to access [21]. User characteristics and expertise can also influence their interaction abilities and preferences. For example, in *AudioFunctions.web*, a cross platform web app for accessing math function graphs through auditory feedback, participants with greater expertise in mathematics found it easier to interact with the system with all interfaces, including traditional mouse/touchpad interaction, while participants with lower math expertise preferred mobile touchscreen access [28].

3) ***Cultural and Linguistic Factors****:* unlike in-situ experiments, which commonly entail a geographical bias, remote studies can sample participants in different linguistic and cultural settings. Usage data analysis by language and localization settings may capture cross-cultural differences, suggesting the need for in-depth app internationalization. For example, in *WordMelodies*, which proposes exercises to support the literacy of children with visual impairments [22], remote data collection unveiled that most popular exercise types and common errors significantly differ between English and Italian users, thus indicating most interesting and important exercises for the two linguistic groups.

B. Usage Behavior Analysis

This analysis is based on the idea that the users often interact with the system in specific use cases, thus following a consistent behavior. By mapping user interaction sequences with the system, such behavior can be captured, and most common patterns can be identified. One methodology we adopted is the use of Natural Language Processing (NLP) techniques for identifying most common interaction sequences [30]. In this approach, each interaction is represented as a word, and sequences of interactions are mapped as n-grams, that is contiguous sequences of n items. We analyze n-grams at different granularity levels.

1) ***Identification of Common Interaction Sequences****:* detecting common short n-grams (with a small n) allows us to capture the interactions that are often executed in sequence [16]. This information can be used for improving the app interface in order to facilitate the execution of typical interaction sequences. For example, in *iMove around*, we detected that some users would open the app, check their current address and close the app immediately, signifying that simply checking their location was one of the common use cases. This lead to providing such information directly on the main app screen.

2) ***User Clustering Based on Interaction Sequences****:* instead, detecting common longer n-gram sequences may be used to identify more complex behaviors and use cases. We used this information to cluster users based on frequency analysis of common behaviors displayed by such sequences [20]. For example, in *iMove around*, the cluster of users accessing information on their surroundings sporadically in short bursts correlates with less frequent screen reader usage, indicating users with low vision who do not need support all the time but mostly need a confirmation of their position from time to time while navigating [21].

C. Longitudinal Usage Analysis

Another interesting analysis approach examines how the participants' behavior and opinions change through time while using the application. For example, in the case of *ReCog* personal object recognizer app, which is designed to support blind people in taking pictures of their personal objects and automatically recognizing them, we conducted a longitudinal test with blind participants [25]. Results revealed that prolonged usage of the system with audio feedback support made the participants able to take better pictures of the target objects, even without audio feedback from the app. Yet, despite an improvement in the overall performance, continuous usage of the system resulted in a decrease of the perceived system usability between the beginning and the end of the study, confirming the well known effect that, once the novelty effect subsides, accrued user experience negatively impacts the perceived usability of a system [31].

V. Advantages and Limitations

A. Participants Recruitment

The main advantage for running remote evaluations is that it **eases the process of recruiting participants**. Not only participants can be recruited worldwide, but they do not need to travel towards the experimental location, which can be difficult, in particular for some people with disabilities. This is also particularly relevant when there are restrictions on the movements of people, like in the case of the COVID-19 pandemic. In our previous paper about *iMove around* [20] we collected information from about 4055 users, a number that would be impossible to involve in in-situ experiments. Another collateral effect of selecting participants worldwide is that **geographical and cultural biases are reduced**. However, note that these biases are not totally eliminated, as, for example, only people that can access the application will participate in the experiment (they need a device, internet access and to be informed about the application itself); also, the application is generally localized only in few languages.

A major limitation of Approach 1 is the **lack of a selection criteria**: participants may have unknown characteristics, or they might not even belong to the target user population. For example, the analysis of remote data collected in the first weeks after the publication of *WordMelodies* [22] unveiled that some users were solely exploring the app rather than working on the exercises. We believe that these users might be teachers or parents of children with visual impairments who first explored the app on their own. Clearly, such data should be pruned before analysis. To this end, it is possible to classify participants ex-post using the collected data [21]. However, this is challenging because the collected data might not be sufficient to clearly profile a participant: while it is possible to identify users with severe visual impairment based on screen reader usage, it may be difficult to identify participants with milder forms of impairment, like low vision users.

B. Longitudinal Studies

Approach 1 naturally supports **longitudinal studies**: following users interaction in a long period of time can provide insights that would not be observable in a single evaluation session. Approach 2 eases longitudinal studies too, however in this case it is important to provide a strong motivation for the participants to keep using the application in time. For example, in *ReCog* personal object recognizer application, the ability to perform a longitudinal study allowed us to observe a learning effect: the participants improved in autonomously taking well-framed pictures of the objects to detect, even without using the supporting audio feedback provided by the system. However, of the ten participants, two stopped using the system early-on and therefore had to be excluded from the longitudinal analysis. Indeed, an aspect to be considered is that participants can **stop using the application**. On one side, this is a limitation, as it reduces the number of participants from which the researcher can collect information. On the other side, however, with Approach 1 (and to a minor extent Approach 2), it is possible, in principle, to study why participants stop using the application, which can yield interesting insights on the application itself. This is however hard to do in practice, because the analysis can only rely on the data collected before the user stops using the application. For example we are currently investigating why users stop using our *MathMelodies* and *WordMelodies*, that support learning of primary school children with visual impairment with a set of exercises. We are trying to figure out whether the exercises are too complex (hence the application is frustrating for the participants) or too simple (hence the application is boring).

C. Engineering Effort

There are two technical constraints with the remote data collection methodology. First, the researcher needs to control which logs to collect and to access the logs themselves; this typically means that the remote experimental methodology can be used with the applications developed by the researcher and not with others (e.g., third party applications already available on the market). Second, this approach can **only** be applied to **software solutions**, already implemented as advanced prototypes, that run on general purposes devices available to many users. To reach the level of a robust advanced prototypes, typically an initial stage of prototyping, possibly involving also in-situ experiments, might be needed. Thus, there is a trade-off in the needed **researcher effort**. On one side, Approach 1 is more scalable: it allows to involve a large number of participants with a small effort by the researcher. However, the engineering process needed to create an application suitable for Approach 1 requires a larger effort than producing an early prototype like those typically used for in-situ experiments. The same reasoning applies, to different extent, to Approaches 2 and 3. In general, the more automated the data collection process is, the less effort is required to collect data, but a larger effort is required to create the application itself.

D. Experimental context

The *Hawthorne effect* [3] is particularly relevant for the evaluation of assistive technologies. Consider for instance to evaluate, with an in-situ experiment, a navigation application for people with visual impairments: the participants can feel safer thanks to the presence of the supervisor (preventing hazards to the participants) but they can be asked to move in an unknown area, which is unusual for many people with visual impairments. More generally, **in-situ experiments often consider a single stereotyped context** without capturing a number of factors, including the lighting conditions or the weather (relevant for applications using computer vision), ambient noise (relevant for applications using audio feedback), or the participant activity like walking or standing (relevant when studying touchscreen gestures, like in typing applications).

Conducting a **remote experiment** (in particular with Approach 1) makes is possible to **evaluate the app in different contexts** that would be impossible to reproduce in in-situ experiments. On the other hand, **it is often difficult**, if not impossible, **to infer the exact context from the logged data**.

VI. Conclusions

Most research in the field of assistive technology present experiments conducted in-situ. There are good reasons for this, in particular these experiments can be conducted with early prototypes (or even without a working prototype, as in the case of Wizard of Oz experiments [32]), and direct participants' observation provides insights. On the other hand, this approach to users' studies also has a number of problems, in particular the limited number of participants that can be involved and the duration of the experiment for each participant. This tutorial paper introduces a different approach, based on the remote collection and following analysis of user interaction data.

With this contribution we are not arguing that remote evaluation can substitute in-situ experiments. Instead, we support that the two approaches can be used at different steps of the research process. In the earlier stage of the research, an advanced prototype is generally not available and direct users' observation is necessary to guide the interactive design process, so in-situ experiments is required. Once the application user interaction has been refined and a more advanced prototype is available, remote evaluation can be used to conduct experiments with a much larger set of participants.

References

[1] A. Sears and V. Hanson, "Representing users in accessibility research," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2011, pp. 2235–2238.

[2] S. Eraslan and C. Bailey, "End-user evaluations," in *Web Accessibility*. Springer, 2019, pp. 185–210.

[3] R. Macefield, "Usability studies and the hawthorne effect," *Journal of usability studies*, vol. 2, no. 3, pp. 145–154, 2007.

[4] A. Steed, F. R. Ortega, A. S. Williams, E. Kruijff, W. Stuerzlinger, A. U. Batmaz, A. S. Won, E. S. Rosenberg, A. L. Simeone, and A. Hayes, "Evaluating immersive experiences during covid-19 and beyond," *Interactions*, vol. 27, no. 4, pp. 62–67, 2020.

[5] H. Panosian, "Testing your app," in *Learn iOS Application Distribution*. Springer, 2017, pp. 147–151.

[6] S. Trewin, D. Marques, and T. Guerreiro, "Usage of subjective scales in accessibility research," in *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, 2015, pp. 59–67.

[7] K. Caine, "Local standards for sample size at chi," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 981–992.

[8] A. Aizpurua, M. Arrue, S. Harper, and M. Vigo, "Are users the gold standard for accessibility evaluation?" in *Proceedings of the 11th Web for All Conference*, 2014, pp. 1–4.

[9] D. Rios, S. Magasi, C. Novak, and M. Harniss, "Conducting accessible research: including people with disabilities in public health, epidemiological, and outcomes studies," *American journal of public health*, vol. 106, no. 12, pp. 2137–2144, 2016.

[10] G. Guest, "Sampling and selecting participants in field research," *Handbook of methods in cultural anthropology*, pp. 215–249, 2014.

[11] D. Ahmetovic, R. Manduchi, J. Coughlan, and S. Mascetti, "Mind your crossings: Mining gis imagery for crosswalk localization," *Transactions on Accessible Computing*, 2017.

[12] J. Kjeldskov and M. B. Skov, "Was it worth the hassle? ten years of mobile hci research discussions on lab and field evaluations," in *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, 2014, pp. 43–52.

[13] N. Dell, V. Vaidyanathan, I. Medhi, E. Cutrell, and W. Thies, "" yours is better!" participant response bias in hci," in *Proceedings of the sigchi conference on human factors in computing systems*, 2012, pp. 1321–1330.

[14] F. Paternò, A. Russino, and C. Santoro, "Remote evaluation of mobile applications," in *International Workshop on Task Models and Diagrams for User Interface Design*. Springer, 2007, pp. 155–169.

[15] K. L. Jensen and L. B. Larsen, "Evaluating the usefulness of mobile services based on captured usage data from longitudinal field trials," in *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, 2007, pp. 675–682.

[16] S. Sultan and P. Mohan, "Transforming usage data into a sustainable mobile health solution," *Electronic Markets*, vol. 23, no. 1, pp. 63–72, 2013.

[17] Y. He, C. Wang, G. Xu, W. Lian, H. Xian, and W. Wang, "Privacy-preserving categorization of mobile applications based on large-scale usage data," *Information Sciences*, vol. 514, pp. 557–570, 2020.

[18] Y. Bhavnani, K. Rodden, L. C. Guarnotta, M. T. Lynn, S. Chizari, and L. Granka, "Understanding mobile phone activities via retrospective review of visualizations of usage data," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2017, pp. 1–10.

[19] J. Froehlich, M. Y. Chen, S. Consolvo, B. Harrison, and J. A. Landay, "Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones," in *Proceedings of the 5th international conference on Mobile systems, applications and services*, 2007, pp. 57–70.

[20] H. Kacorri, S. Mascetti, A. Gerino, D. Ahmetovic, H. Takagi, and C. Asakawa, "Supporting orientation of people with visual impairment: Analysis of large scale usage data," in *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. ACM, 2016.

[21] H. Kacorri, S. Mascetti, A. Gerino, D. Ahmetovic, V. Alampi, H. Takagi, and C. Asakawa, "Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data," *Transactions on Accessible Computing*, 2018.

[22] S. Mascetti, G. Leontini, C. Bernareggi, and D. Ahmetovic, "Word-melodies: supporting children with visual impairment in learning literacy," in *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. ACM, 2019.

[23] N. Cantù, M. Ducci, D. Ahmetovic, C. Bernareggi, and S. Mascetti, "Mathmelodies 2: a mobile assistive application for people with visual impairments developed with react native," in *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. ACM, 2018.

[24] D. Ahmetovic, S. Mascetti, C. Bernareggi, J. a. Guerreiro, U. Oh, and C. Asakawa, "Deep learning compensation of rotation errors during navigation assistance for people with visual impairments or blindness," *Transactions on Accessible Computing*, 2019.

[25] D. Ahmetovic, D. Sato, U. Oh, T. Ishihara, K. Kitani, and C. Asakawa, "Recog: Supporting blind people in recognizing personal objects," in *ACM Human Factors in Computing Systems (CHI)*. ACM, 2020.

[26] C. Gleason, D. Ahmetovic, S. Savage, C. Toxtli, C. Posthuma, C. Asakawa, K. M. Kitani, and J. P. Bigham, "Crowdsourcing the installation and maintenance of indoor localization infrastructure to support blind navigation," *PACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) - ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2018.

[27] A. Gerino, L. Picinali, C. Bernareggi, N. Alabastro, and S. Mascetti, "Towards large scale evaluation of novel sonification techniques for non visual shape exploration," in *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, 2015, pp. 13–21.

[28] D. Ahmetovic, C. Bernareggi, J. a. Guerreiro, S. Mascetti, and A. Capietto, "Audiofunctions.web: Multimodal exploration of mathematical function graphs," in *International Cross-Disciplinary Conference on Web Accessibility (W4A)*, 2019.

[29] R. Fielding, "Representational state transfer," *Architectural Styles and the Design of Netowork-based Software Architecture*, pp. 76–85, 2000.

[30] C. Bertero, M. Roy, C. Sauvanaud, and G. Trédan, "Experience report: Log mining using natural language processing and application to anomaly detection," in *2017 IEEE 28th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 2017, pp. 351–360.

[31] P. Kortum and M. Johnson, "The relationship between levels of user experience with a product and perceived system usability," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2013, pp. 197–201.

[32] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies: why and how," in *Proceedings of the 1st international conference on Intelligent user interfaces*, 1993, pp. 193–200.