Anchored Audio Sampling

A Seamless Method for Exploring Children's Thoughts During Deployment Studies

Alexis Hiniker
University of Washington
Seattle, Washington, USA
alexisr@uw.edu

Jon E. Froehlich
University of Washington
Seattle, Washington, USA
jonf@cs.washington.edu

Mingrui Zhang
University of Washington
Seattle, Washington, USA
mingrui@uw.edu

Erin Beneteau
University of Washington
Seattle, Washington, USA
ebenet@uw.edu

ABSTRACT

Many traditional HCI methods, such as surveys and interviews, are of limited value when working with preschoolers. In this paper, we present *anchored audio sampling* (AAS), a remote data collection technique for extracting qualitative audio samples during field deployments with young children. AAS offers a developmentally sensitive way of understanding how children make sense of technology and situates their use in the larger context of daily life. AAS is defined by an *anchor event*, around which audio is collected. A sliding window surrounding this anchor captures both antecedent and ensuing recording, providing the researcher insight into the activities that led up to the event of interest as well as those that followed. We present themes from three deployments that leverage this technique. Based on our experiences using AAS, we have also developed a reusable open-source library for embedding AAS into any Android application.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**.

KEYWORDS

Study methods, field deployments, early childhood, CCI, data collection tools, context-aware ESM

ACM Reference Format:
Alexis Hiniker, Jon E. Froehlich, Mingrui Zhang, and Erin Beneteau. 2019. Anchored Audio Sampling: A Seamless Method for Exploring Children's Thoughts During Deployment Studies. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300238

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300238

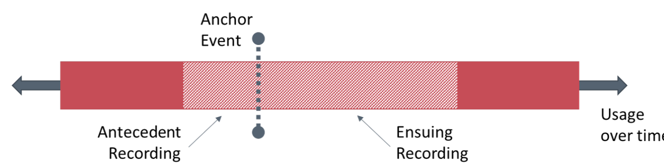

Figure 1: Timeline of an AAS data collection event. The researcher defines events of interest, which become the anchor points. Recording occurs during a sliding window surrounding the anchor event.

1 INTRODUCTION

Collecting rich, experiential data from young children about their perspectives on novel technologies can be challenging. Young children are more likely than adults to be unaware of what they know [28] and may struggle to engage meaningfully with survey or interview questions [22]. Yet, children are avid users of technology [3, 41] and experts of their own experiences [20], and the research community has increasingly recognized the importance of involving children in the process of creating and evaluating novel designs and systems [22, 24, 27]. Prior work has shown that parents, teachers, and other adults are often poor proxies for understanding children's experiences [20], meaning that these methodological limitations cannot be addressed simply by shifting the data collection burden to a trusted adult. Thus, to most effectively generate valid and valuable insights about technologies for children, researchers must have mechanisms for capturing children's reactions directly.

Furthermore, understanding technology in its natural context of use is valuable with users of any age [58]. A number of well-honed techniques allow researchers to collect data in the wild, including diary studies (*e.g.*, [43, 62]), experience sampling (*e.g.*, [45, 56]), and passive data logging (*e.g.*, [54, 55]). Although all of these techniques offer value in understanding young children's experiences, they also have limitations: pre-literate children cannot engage with ESM or diary study procedures in their traditional formats, and passive sensing alone cannot capture qualitative data about children's perspectives and experiences. Thus, although many field studies with children have generated valuable design insights [39], there remains a need for methodological advances that target this population and support remote, qualitative data collection at scale [22].

In this paper, we describe *anchored audio sampling* (AAS), an event-triggered technique for collecting rich, qualitative data from users as they engage with a technology. At its core, AAS involves using a microphone to collect snippets of audio about a participant's real-time experiences. These audio recordings are collected in response to specific events of interest, which we call *anchor events*. To support the research team in understanding the factors leading up to these events, AAS uses a parameterizable back buffer to collect audio that came before the anchor event, a mechanism we call *antecedent recording*.

To examine AAS and synthesize its key benefits and drawbacks, we re-examine three prior AAS-based field deployments: 1) a study of a video player designed to ease preschoolers' transitions away from the device when videos end, 2) an evaluation of an app for preschoolers designed to improve executive function, and 3) an evaluation of the *Amazon Echo Dot* smart speaker. Through an inductive-deductive secondary analysis of the audio recordings and other data collected across studies, we extracted key themes, such as the usefulness of AAS in situating children's experiences, the value of scoping the dataset to moments of interest, and the tension between raising participants' awareness of recordings and altering their behaviors.

Given that this technique creates a low barrier to accessing intimate moments of family life, it also requires sociotechnical structures to ensure fully informed ongoing consent [2] about what is being captured and when. In addition to describing the parameters for collecting data, we also describe our techniques for notifying families about audio capture and mechanisms for giving them control of their audio data. Prior work calls for greater transparency in technologies that record children's speech [46], and a central motivation for formalizing this method is to tightly couple the process of capturing audio from children with the process of considering how best to manage this data.

In addition to providing a methodological contribution for collecting contextualized technology usage data from young children at scale, we also contribute an open-source AAS library that can be plugged into any Android application. This library includes recording features, an optional back buffer, multi-modal UI for notifying participants about recordings, a pluggable interface that allows participants to review and manage their own data, and a number of other customizable features. In doing so, we hope to make it easier for the HCI and child-computer interaction research communities to make use of AAS and to do so with maximal consideration for the privacy and comfort of participants.

2 RELATED WORK

2.1 Evaluating Technologies in the Wild

In-the-wild evaluations are central to the field of HCI [49, 58], as contextual factors are a key component of the way people experience a technology. As direct observation is not always possible or desirable during field deployments, a number of methods support researchers in capturing user data in unsupervised contexts [50]. For example, remote usability testing was developed in the 1990s, and within a decade became popular in both research studies and commercial software development [8]. Although in some instances, remote usability testing is synchronous (*i.e.*, the evaluation is monitored in real-time), it is most often asynchronous, with researchers or product teams separated from users in both space and time [7].

Historically, the techniques for conducting asynchronous remote usability testing have ranged from deploying online questionnaires (*e.g.*, [48, 57]) to logging interactions (*e.g.*, [21]). Today, passively conducting experimental A/B tests on users' interactions with apps and websites is common practice [26]. Although remote usability testing is widespread, and many tools exist to support it [6], a recent literature review reports that this rarely includes qualitative measures of users' experiences [59].

Similarly, the experience sampling method (ESM) [44] (also known as ecological momentary assessment [63]) is a well-established method of collecting *in situ* self-report data from study participants at times specified by the research team. Prior work explains that ESM is well-suited to in-the-wild evaluations of novel technologies [13]. Work in HCI has not only embraced ESM as a common study procedure, it has also innovated methodologically by inventing novel implementations for supporting ESM and enabling context-driven (rather than random) sampling. For example, Carter and colleagues created Momento, a system for integrating log data with qualitative experience sampling [11], and Froehlich and colleagues created MyExperience, a tool for enabling context-aware ESM data collection via passive sensing [25]. A number of other studies have implemented passive sensing systems to give researchers insight into participants' real-world experiences (*e.g.*, [4, 16, 17, 23, 61, 66]). AAS is similar to context-aware ESM: it collects data in the moment in response to environmental triggers. However, AAS responses are richer than most ESM samples and they do not depend on self-report, enabling remote, context-specific, qualitative data collection at scale.

2.2 *In Situ* Audio Capture

Audio recording is a staple of usability research and routinely employed to capture interview and observational data during in-person studies. It is also useful in unsupervised contexts, and for example, Hertzum and colleagues report that traditional think-aloud protocols produce a higher density of relevant verbalizations when they are unmoderated and conducted outside the lab [32].

A number of prior studies have sought to establish techniques for integrating audio capture into deployments in the wild. Palen and Salzman expand on the traditional diary study format with a method that enables participants to document their experiences via voicemail [51]. Rachuri and colleagues developed a platform for conducting social psychology experiments that includes processing speech samples from participants to predict the individual's mood [53]. SociableSense captures input from a device microphone to use the presence of audio information as a proxy for socialization [52]. And NoiseSPY leverages mobile phone microphones to detect noise levels across an entire city, thereby using audio sampling to understand the human experience at a macro level [42]. Our work continues this tradition of leveraging samples of audio data in unsupervised contexts to learn more about users, their broader context, and their experiences with technology.

2.3 Methods for Designing for Children

Though AAS can be of use in studies with participants of any age, we developed this technique in response to a need for methods that are effective in deployment studies with young children. Understanding children's experiences and perspectives requires soliciting and listening to their voices directly [29]. However, many of the most commonly used methods for soliciting user input in HCI studies are less effective with young children, who may be unable to read [68], unaware of what they know [28], or struggle to articulate ideas with sufficient detail and context such that others can understand [12].

As a result, a number of methodological innovations have been designed for the explicit purpose of conducting research with children. Perhaps the most well-known research method in this space is cooperative inquiry [19], pioneered by Druin and colleagues. This participatory approach builds on contextual inquiry [35] and situated action [64] to foreground children's voices during the design process. A number of other methodological studies have built on this work to explore how research methods can elevate children to be full design partners [20] during the development of new technologies. For example, Farber and colleagues extend these techniques to preschoolers [24], and Yip and colleagues examine what full partnership between adults and children looks like in practice [69].

Other researchers have innovated methods for collecting contextualized details about children's experiences in the wild, and a literature review demonstrates that field studies are one of the most common means of evaluating children's technologies [39]. For example, Einarsdottir and colleagues collect details about children's situated experience by asking them to take photos throughout the day with a disposable camera [22]. Jones et al. developed a system enabling children to proactively share thoughts with a toy with recording functionality [40], and the photovoice method [65] is frequently used with older children as an elicitation technique for understanding their experiences with and relationship to their environment [5]. Yet the majority of work with children continues to be conducted with those age 6 and older [38]. Our work offers a methodological approach that can be used with children of any age, and specifically supports younger users who perspectives are more likely to be overlooked.

3 ANCHORED AUDIO SAMPLING

AAS entails passively recording snippets of ambient audio in response to a specific event of interest, optionally including back-buffered audio that preceded the event. Audio recording occurs seamlessly without the need for manual intervention from the end user. Here, we describe the four core components of the method: anchor events, antecedent and ensuing recording, real-time feedback, and data storage and propagation.

3.1 Anchor Events

The data that is collected during an AAS study is defined by *Anchor events*, occurrences that the researcher has selected as triggers for audio recording. Conceptually, this term is an abstraction and simply refers to any arbitrary event the researcher has predetermined to be relevant to the research question at hand. Thus, anchor events might refer to any number of user behaviors, activities, or states, such as:

- The user interacting with a particular feature
- The user saying a particular word
- The user entering a particular usage phase, such as starting a level in a game or completing an activity
- The user repeatedly failing to perform an action
- The user exhibiting a particular physiological state, such as a high heart rate reported by a passive sensor
- The user entering a particular context, such as a specific location or time of day

These moments define the data collection process: the dataset that results from an AAS study is a collection of audio snippets that surround anchor events. Given that anchor events can range from discrete user events (*e.g.*, a user clicks a button) to probabilistic predictions (*e.g.*, a "stress inference engine" predicts that a user is stressed), the accuracy of AAS, much like the accuracy of ESM, is a function of the detection techniques chosen by the research team. Anchor events must be able to be operationalized and detected passively to enable seamless audio recording.

3.2 Antecedent and Ensuing Recordings

AAS supports two forms of data collection in response to an anchor event. *Antecedent recording* preserves audio data that was collected before the event was detected, and *ensuing recording* begins capturing audio once the event is detected. For example, a study that seeks to understand reasons why a child stops playing a game could use app closure as an anchor event and collect an antecedent recording to understand what led up to the child closing the app (such as signs of frustration or boredom).

The length of both antecedent and ensuing recording periods is set by the research team *a priori*. In many cases, the recording window may be best specified by time (*e.g.*, a minute of audio data before the anchor event and a minute after). However, in some instances the recording window that surrounds the anchor event may be better specified by other contextual attributes. For example, a researcher studying a child's experience during high-intensity moments of game play might use a high heart rate as an anchor event. The recording window could then be delimited by the most recent point at which the participant's heart rate was within 10% of average resting rate and the next time it returns to this baseline. Or a deployment study of a drawing app interested in understanding when and why a child chooses to share drawings with others might collect audio data anchored around the point when the child presses a "publish" button and extending back to the point when the child began drawing.

To enable antecedent recording, the system continuously collects audio data, deleting unneeded snippets after it becomes clear they will not fall within an anchored recording window (*e.g.*, each time one minute passes, each time the child's heart rate drops back into the resting range, or each time the child begins a new drawing). Thus, when the start of the recording window is specified by context, the researcher should consider the cadence of these context-specific windows and whether this might lead to extensive (and therefore potentially invasive or logistically impractical) recording durations. If so, it may also be necessary to set an upper-bound on the length of the antecedent recording and to delete snippets based on time as well as context.

3.3 Real-time Feedback for Participants

AAS data is rich and personal, yet it is not self-reported, making it harder for participants to self-censor what they disclose, both a strength and limitation of the method. Ensuring ongoing informed consent throughout an AAS study is challenging, as participants may not readily recall in real-time when they are being recorded or when antecedent recordings are preserved. This is particularly challenging in studies with young children, where the user of the system may be a different person from the one who consents to participate. As a result, providing feedback to notify participants of anchor events as they occur and remind them in-the-moment that they are being recorded is a central component of the method.

This feedback can be in the form of on-screen UI, if the participant is likely to be looking at the system when using it. If not, other techniques, such as a specific sound, an audio announcement, or haptic vibration can all serve the same purpose. If an adult is unlikely to be sharing the experience with the child, researchers might consider sending a push notification to a parent's phone or providing another form of feedback that can alert a trusted adult that recording is active.

3.4 Storage and Propagation

To provide increased support for ongoing, informed consent and mitigate the risk of capturing audio that families would prefer not to share, an AAS study also requires the researcher to determine whether and how to give participants *post hoc* control over their data. In thinking about *post hoc* control, the researcher should address two questions.

First, will participants have the opportunity to review and potentially delete recordings? Giving participants access to all recordings and the ability to review and delete their own data is particularly important in light of the fact that antecedent recording, by definition, is captured before an anchor event occurs, making it less likely that participants are aware of audio capture during this period. It is also important given that a child may be engaging with the system alone, but an adult is likely to be better positioned to identify and redact sensitive data. Through our experiences deploying AAS studies (see section 5), we have found that giving participants a mechanism for deleting data has not undermined our ability to collect a robust data set and that participants value and make use of this option.

Second, when do researchers get access to recordings? This might happen automatically and immediately, automatically but after a delay, or manually when the participant chooses to share them. This introduces trade-offs between maximizing participant agency and tolerating potential data loss. It also demands consideration of the user burden, as

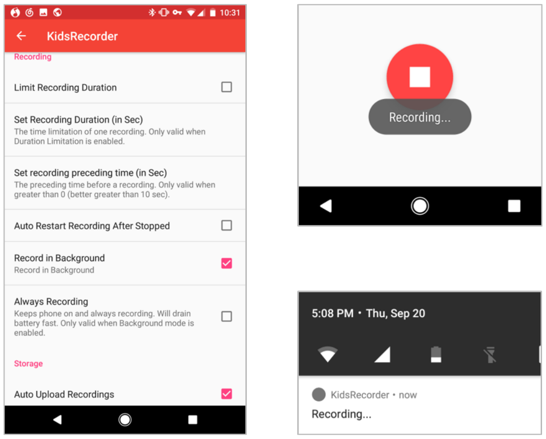

Figure 2: Left: Default settings menu. All settings can be adjusted via API or UI, giving access to both the research team and the participant. The developer can choose to customize the settings menu to hide options from the participant. Top Right: Default UI for recording; a fleeting toast notification appears on screen when an anchor event occurs. Bottom Right: Default UI for recording; a persistent notification appears in the notification shade.

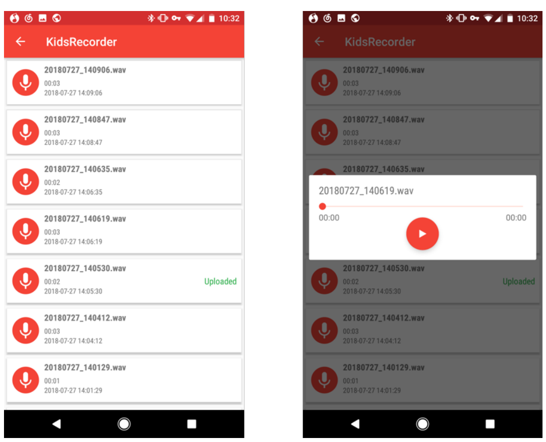

Figure 3: Default UI for reviewing recordings. Left: the participant can view their complete data set. Right: Each recording can be reviewed.

reviewing and sharing data requires effort on the part of the participant.

4 OPEN-SOURCE AAS ANDROID LIBRARY

To support broad use of AAS in future studies and to encourage researchers to provide transparency and control to users, we have designed and released a general-purpose, open-source AAS library [1] for Android. Our library addresses the four components of AAS in the following ways:

[1]https://github.com/uelab/KidsRecorder

- *Anchor events:* These must be defined and detected by the research developer and are not directly supported by our library; upon detecting an anchor event, the research developer should call our library.
- *Antecedent and ensuing recording:* Our library enables antecedent recording via a continuous back buffer. It stitches together antecedent and ensuing recordings, parameterized by the research developer, storing a single recording for each anchor event.
- *Real-time feedback:* Our library provides multi-modal, redundant, parameterizable feedback to signal anchor events and inform the user when recordings are being collected.
- *Storage and propagation:* Our library provides a UI view allowing the user to review, delete, or share recordings.

Initiation. To embed these components in a deployment system, the research developer links to our library and then creates singleton RecordingManager and DataManager objects. The recordingManager object enables the research developer to control recording activities. As RecordingManager extends Service, the research developer should then start and bind to a service with a RecordingManager Intent. The dataManager object gives the research developer access to stored recordings. Optionally, the research developer can then configure parameters of the recording experience. For example, she might choose to set a flag to allow recording only when the deployment app is the active app:

```
recordingMangager.setAlwaysRunning(false);
```

Recording. Once the recording parameters have been configured, the deployment system is ready to collect AAS samples. When the deployment system detects an anchor event, it should request an AAS sample from the recordingManager using the StartRecording API:

```
recordingManager.StartRecording(
    metadataString,
    antecedentDuration,
    ensuingDuration);
```

Depending on its parameters, this call initiates an ensuing recording and/or preserves an antecedent recording, ultimately stitching the two together into a single sample. The API also allows for an optional metadata string to attach to the file name. When recording, the library uses the built-in device microphone, accessed via Android's MediaRecorder APIs in SDK v23+.

Real-time UI Feedback. The library provides default UI to notify participants in the moment that recording is active, including an audio chime, a haptic vibration, a persistent notification in the notification bar, and an on-screen toast notification (see Figure 2, right). The developer can adjust these defaults through the RecordingManager API.

We also provide a default settings menu, which can be used both by the research developer to manually set defaults and by the participant to adjust UI feedback to their comfort level (see Figure 2, left). The research developer can adjust what is exposed in the settings menu, allowing the participant to control some settings while obscuring those that the participant should not adjust.

Storing and Propagating Recordings. Recordings are preserved locally as .wav files and, by default, named according to the timestamp of the anchor event. To reduce the technical burden to giving participants full control over their own data, we provide sample UI for reviewing all recordings (see Figure 3 3). This enables participants to preview and vet data locally before sharing it with the research team.

By default, recordings are not automatically propagated to the research team. However, researchers can opt to adjust this through the dataManager:

```
dataManager.setAutoUpload(true);
```

We provide default integration with the *AWS S3* cloud storage system [1] and guidance for propagating audio recordings to an *AWS* account (which the research team must create). However, our library can be used with any cloud platform or other backend data-storage system.

5 CASE STUDIES

Here, we describe three prior AAS deployments (see Table 1) as case studies that present both instances of using the method and motivation for its evolution over time.

5.1 Overview of Case Studies

Study 1: Preschoolers' Video Transitions. In its first instantiation, we embedded AAS into a video player for preschoolers [33]. The purpose of the study was to understand children's transition experiences when a video playlist ends, asking specifically: How do children react when it is time to stop watching videos? Can the design of a video player influence this experience?

To answer these questions, we created a custom video player as a front-end wrapper for *YouTube*. We included features to enable the user to: build a playlist, select a non-digital activity to engage in after the playlist ends, watch the videos in the playlist, and walk through a transition experience, during which the system reminds the user of their planned next activity (see Figure 4, left).

We conducted a three-week in-home deployment with 24 preschoolers (ages 3-5) and their families, during which time we asked them to use our video player once each day. We automatically varied elements of the transition experience each week to examine how children's behaviors changed with the design. To study children's responses during the point of transition, we captured one minute of antecedent audio and two minutes of ensuing audio, anchored around the end of the playlist. For complete study details, see [33].

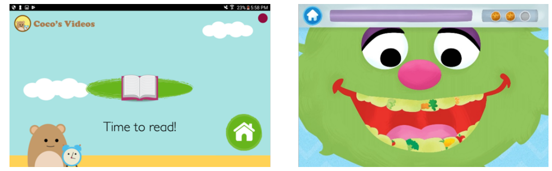

Figure 4: Screenshots from case study apps. Left: In Study 1, a transition screen is displayed when a playlist ends, reminding the child of the next activity she planned for herself (here, reading). Right: In the tooth-brushing minigame of *Cookie Monster's Challenge*, the player scrubs the monster's teeth clean.

Study 2: A Tablet App to Train Executive Function. In its second instantiation, we used AAS as part of a mixed-methods evaluation of *Cookie Monster's Challenge*, a tablet game for preschoolers designed to train executive function (*i.e.*, a composite skill that combines working memory, inhibitory control, and cognitive flexibility, known to be a strong predictor of a number of outcome measures later in life). The purpose of this study was to determine whether *Cookie Monster's Challenge* leads to increases in executive function, and if so, what behaviors and strategies children use to engage their executive function as they play the game. We recruited 37 preschoolers to play either *Cookie Monster's Challenge* or an active control together with a parent at home daily for two weeks. We measured executive function in the lab before and after the deployment, and we collected audio during the deployment period. As there was no specific *a priori* interaction event of interest, we collected five minutes of ensuing audio each time the app was opened and again at random times while it was the active app.

Study 3: Families' Use of an Amazon Echo Dot. In a third study, we deployed an *Amazon Echo Dot* smart speaker in the homes of ten low- and middle-income families. We used AAS to collect audio samples on a separate data collection tablet any time a family member (of any age) engaged with the *Echo Dot*, anchored around the use of the wake word "Alexa" and collecting one minute of antecedent audio and two minutes of ensuing audio. All families had at least one child between the ages of 4 and 17, and some had younger siblings. We recorded interactions with the *Echo Dot* for a period of four weeks; for complete study details, see [10].

5.2 Case Study Analysis Methods

We performed a secondary analysis of the original AAS datasets, gathering transcripts of all audio samples collected in each case study (together, comprising more than 1000

Table 1: Overview of Case Studies

	Research Questions	N	Age (yrs)	Anchor(s)	Rec. Duration (min)		Feedback	Propagation
					Antecedent	Ensuing		
1	How do children respond to the transition away from a screen when a video playlist ends?	24	3–5	Video playlist ending	1	2	Recording icon on screen	Automatic, immediate
2	Does children's executive function improve after playing the game Cookie Monster's Challenge, and if so, how?	38	3–5	Game start At random during active use	N/A	5	Recording icon on screen	Automatic, immediate
3	How do low- and middle-income families use smart speakers at home?	36	<1 – >68	Use of the wake word, "Alexa"	1.5	2.5	Written text on screen, audio cue	Automatic, delayed

recordings), as well as exit interview transcripts from Studies 2 and 3 (in Study 1, participants only completed an exit survey, which did not include questions about the data collection experience). We also consulted with research assistants who participated in the consent process in each study, asking them to recall insights about families' reactions to the planned procedures during recruiting, data collection, and debriefing. Finally, we revisited the manuscripts and analysis produced to-date from each study, reviewing the analysis activities, labor, data set size, and findings from each project.

We performed an initial open-coding of this data using an inductive-deductive approach [14] to identify commonalities (both positive and negative) that spanned multiple studies. To do so, the research team re-read all audio transcripts, as well as the supporting study notes and interview transcripts, keeping a list of potential themes. After collaboratively discussing themes, the team then iteratively re-read samples and notes, extracting representative quotes and descriptions for each theme according to the "vivid exhibits" lens [36].

We performed subsequent coding passes to refine and consolidate themes. Themes that were well represented with numerous examples across studies were included here, illustrated with exhibits from the larger collection.

5.3 Case Study Findings: Benefits and Challenges of AAS

Situated Data. Across all three case studies, we routinely found that we were able to capture situated data about participants' use of the deployment system. For example, in the following exchange from Study 2, we hear a parent and child's shared experience as they encounter a tooth-brushing mini-game that requires the player to quickly and carefully scrub a monster's food-filled teeth (Figure 4, right):

> Cookie Monster: (announcing the start of the tooth-brushing game) "*Brush teeth!*"
> Dad: "*Ok. Brush brush brush.*"
> System audio: (Brushing noises)
> Dad: "*Brush brush brush; get the bottom, get the bottom!*"
> Cookie Monster: "*Brush teeth on top and bottom.*"
> Dad: "*C'mon, get the bottom! Good job! Oh you did a good job!*"
> Child: "*Did it!*"
> System audio: (Success chime)

In this study, we ultimately found that children see gains in executive function after playing this game, and that this relationship is mediated by the way in which parents support the child's engagement with the system. Our audio samples revealed not only whether a parent was helping a child, but also demonstrated the specific support strategies that parents provided, such as breaking the task in to smaller parts, by saying things like "*get the bottom*," or encouraging the child to perform the task quickly by repeating the same instruction ("*brush brush*") many times in rapid succession. This close examination of the collaboration between parent and child allowed us to examine the mechanisms by which children see gains in executive function.

Similarly, in Study 1, we found that children often expressed autonomy about turning off the device and switching to their planned next activity. For example, in one instance we heard the playlist end and the system remind the child that she had planned to play outside. We heard the child then announce that it was time to go outside and that she was searching for her boots. In another instance, a child participant announced that it was time to be done watching videos and then asked his mother to put on music so he could dance.

Through audio samples targeting moments of transition, we saw that children took ownership of transitions, extended them to include the rest of the family, and engaged with their surrounding environment (*e.g.*, searching for boots and requesting music) to proactively bring about the transition they had planned using our system.

Antecedent Recording. We consistently saw that antecedent recording augmented the value of AAS by enabling the research team to collect data before it was known to be of interest. For example, in Study 3, as we looked at families' interactions with the *Echo Dot*, antecedent recording enabled us to go beyond examining what families asked the Echo to understand *why* they asked in the first place. For example, in one snippet, we hear:

> Child: "*Alexa, pick a number between 3 and 5.*"
> Echo Dot: "*Your random number between 3 and 5 is 5.*"

It is only because of antecedent recording that we further know the child and his mother had been planning to go to a happy-hour event and trying to decide when to leave the house. By listening to the human conversation that came before, we see that the child is using the *Echo Dot* as a whimsical way to determine their departure time, selecting among 3 pm, 4 pm or 5 pm. Although AAS can capture contextualized usage data without antecedent recording, this feature enhances the researcher's ability to examine the causal mechanisms behind families' usage decisions.

Similarly, in Study 1, antecedent recording enabled us to understand how children prepared for transitions, making announcements to family members like, "*We have one minute left*" and demonstrated how they ritualized the transition supports provided by the system, saying things like, "*I want to see what she [the system] is going to say*". It showed us how parents anticipated and prepared children for transitions saying things like, "*This is going to be done in a minute; wanna go to the grocery store?*" Listening to children prepare for the event of interest, share it with others, and experience proactive support from parents gave us a more robust view of the larger sociotechnical system in which these moments of use were embedded.

Scoped Data Collection. A further benefit of AAS is its ability to scope the data space. One challenge of working with rich, qualitative data is the labor involved in performing methodical analysis [47]. And prior work has shown that passive sensing systems can collect so much data that it can be difficult to find the signal in the noise [60]. Much like context-aware experience sampling [37], AAS intentionally narrows the focus of the investigation by collecting data around a common trigger or triggers.

For example, in Study 1, 24 children used our app at home at least once a day for three weeks. In doing so, they collectively watched 2,452 videos and spent approximately 500 hours with our system. By aligning our anchor point with our core research question, we were able to surgically target our analysis to the moments of use that directly address our core concern (children's transitions). By setting the child's transition point as the anchor event and capturing three minutes of surrounding audio, we sliced out the 14 hours and 36 minutes (3% of the total usage time) that were most relevant to our research question. In Study 3, we similarly analyzed 14.5 hours of audio recordings, drawn from a month-long deployment in ten homes, less than 1% of the total time the system was deployed.

Intrusiveness of Recordings. Even before beginning a study, we found that recruiting participants to AAS deployments was more challenging than with other field deployments. The research assistants who conducted these studies reported that potential participants often brought up questions about audio recording during the enrollment or consent processes, and some potential participants ultimately elected not to participate in these studies for this reason. Although this suggests that potential participants felt informed to make decisions they feel good about, it underscores the fact that these procedures have the potential to cause discomfort, the fact that these procedures come with an added recruiting burden, and the fact that study samples will be biased to only include families who are comfortable with these procedures.

In practice, we found that our case studies did capture intimate moments of family life. For example, we hear the following in Study 1:

> System audio: (A song is playing, then ends)
> Child (to the tablet): "*Bye.*"
> System audio: "*Now it's time to sleep. Are you ready to sleep?*"
> Child: "*Yeah!*"
> Child: "*Mommy, I am ready to sleep.*"
> Mom: "*Good girl. Do you like the timer?*"
> Child: "*Yeah, I like it.*"
> Mom: "*K. I love you.*"

This example characterizes the way in which AAS intermingles user feedback and close, personal moments of family life–here a sweet exchange between parent and child as the child goes to bed. Both the value and challenge of AAS is its ability to give the research team this close view of users' experiences.

Participant Awareness of Recordings. One question that arose in our analysis of all three studies was whether participants' awareness of recordings has the potential to influence behavior. Although rare, we heard explicit comments about active recording, suggesting participants were aware of data collection in the moment at least some of the time. For example, in Study 3 we heard:

> Child: "*Alexa, your jokes are really corny.*"
> Alexa: "*Sorry, I don't know that one.*"
> Child: "*Alexa, get a brain, idiot.*"
> Alexa: "*Hmm, I don't know that one.*"
> Adult: "*Just a reminder that they can hear you.*"
> Child: "*Oh yeah.*"

Here, the parent reminds the child about AAS data collection as the child hurls abuse at the device, suggesting not only that participants are aware that recording is happening, but that they might censor socially unacceptable behavior in response. In another instance, a family in Study 1 explicitly chooses to talk to research team through the recording mechanism:

> Dad (to child): "*Well, what do you think about that show that was just on? . . . Do you have anything you want to say?*"
> Child: "*Play!*"
> Dad: "*What do you have to say about this whole experience? With the Samsung tablet?*"

Here, the father prompts his child to share his thoughts about the experience. In some sense, this can be seen as opportunity for participants to communicate with researchers. However, it also confirms that participants are aware of being recorded in the moment.

Systematic Blind Spots. Though we collected rich data sets in all three case studies, we consistently captured an incomplete picture of the usage context. AAS lacks the visual data that in-person observation or video recording might provide, and in our samples it is at times unclear who is speaking or to what they are referring. At times, a child shouts to someone in the distance and only one side of the conversation is captured. In other instances, two children play with a device together, and it is unclear who is touching the screen. We have found audio to be a useful but incomplete representation of participants' usage and context.

Without additional augmentation, AAS also lacks integration with system logging or screen recording, and prior work has shown the value of combining qualitative and quantitative data during deployments in the wild [11]. In our case studies, we manually triangulated with other data sources; for example, in Study 1, we used timestamps to link the videos children chose to watch as reported by in-app logging with the audio recordings captured when the videos ended.

5.4 Evolution of AAS

As we conducted these studies, we iterated on our approach in three key ways. First, we incrementally formalized participants' ability to manage their data and delete recordings. In Studies 1 and 2, we gave participants the option to contact us to manually delete any recordings they retroactively decided they would prefer not to share. No participants chose to do so, leaving us wondering retrospectively whether this choice meant they were comfortable with the audio they shared, or if the option to request manual deletion was too cumbersome or exposing to be accessible.

Thus, in Study 3, we made this option more accessible by delaying propagation of recordings for 10 minutes and providing UI allowing participants to delete the last 10 minutes of data at any time, redacting this information before it ever reached the research team. Three of 10 families who participated in Study 3 chose to delete data, each deleting exactly one 10-minute block during the month-long deployment. However, during exit interviews, one participant said they once meant to use the delete function but forgot; another said that they once decided not to use the delete button but later wished they had. Based on these suggestions that participants would value additional control, we made further iterations to the data management UI. In our generalized library, we provide participants with full access to review and manage their data, and we provide researchers with the ability to adjust this level of access to meet the needs of their study.

Second, we evolved the level of real-time feedback we provide during active recording. In Study 1 and Study 2, we notified participants of active recording with a red recording indicator on screen. In Study 3, our research question explored a new modality (voice input), making our visual UI indicator insufficient, as participants did not need to look at the deployment tablet when interacting with the *Echo Dot*. Thus, we added audio feedback announcing anchor events.

In extending this to a generalized library, we included other redundant indicators in multiple modalities (visual, audio, and haptic feedback, with both ephemeral and persistent notifications). Through this diverse set of options, we hope to support participant awareness across a broader range of scenarios, including deployments with visually dense UI, deployments with competing audio from the system, and deployments with a diverse range of users including those who are blind or low-vision or otherwise lack access to some indicators.

Third, we saw in Study 1 and Study 2 that participants occasionally chose to treat recording windows as opportunities to speak to the research team (such as the example from Study 1 where a father asks his child if there is anything she would like to say about her experience). As a result, in Study 3, we added the ability for participants to trigger recording on demand, leveraging the infrastructure we put in place for anchor events as a means for participants to leave voice memos and in-the-moment reflections. Across the duration of the study, six of the ten families chose to take advantage of this optional feature, leaving 12 voice memos for the research team about their real-time experiences with the system.

6 DISCUSSION

6.1 Contributions of AAS

Across all of our case studies, we found that we were able to collect rich, situated data from children and families that gave us direct and nuanced insight into their usage experience. We saw that when using our video player, children autonomously announced to their parents when their videos ended and, for example, proactively began searching for boots to go outside. We saw that parents scaffolded challenging executive function tasks for their children, breaking them down into smaller steps. And we saw that children's interactions with the *Amazon Echo Dot* that might appear to be instrumental (*e.g.*, choosing a random number between three and five) are often undergirded by a whimsical intent (giving Alexa the power to choose the departure time for a party).

In its current form, AAS offers three core contributions. Although it shares some of these strengths with existing *in situ* methods, together its collective features offer new opportunities. First, AAS allows researchers to collect qualitative data without requiring self-report from users, reducing the participation burden and potentially reducing the extent to which participants alter their behaviors. Second, by using anchor events to trigger data collection, AAS reproduces the real-time, event-based sampling that characterizes context-aware ESM. We have found this to be particularly valuable in our qualitative case studies, as this approach has allowed us to scope what would otherwise be an enormous expanse of potential data into a corpus tailored with precision to our research question.

Finally, through antecedent recording, we support retroactive data capture that is not typically part of ESM procedures. Prior work has shown that retrospective audio data can be useful; for example, Experience Buffers [31] enable caregivers to retroactively capture audio about experiences they later realize are meaningful, the Personal Audio Loop supports individuals in retroactively preserving their own past audio content [30], and SenseCam facilitates recalling past memories [34]. A key contribution of this work is blending back-buffered audio capture with context-aware ESM.

6.2 Limits to AAS

Yet, we also encountered limits to the benefits of AAS. Much like context-aware ESM, the accuracy of AAS data collection is only as good as the accuracy of the sensing techniques it is built on. When our anchor event was discrete (*i.e.*, the end of a video playlist in Study 1), we could expect virtually perfect accuracy, but when our anchor event was probabilistic (*i.e.*, detection of the word "Alexa") we were dependent on the underlying inference engine, here, the particular speech recognition model we chose.

As the state of the art in passive sensing and prediction evolves, so too will the potential of AAS. Prior work has shown imperfect but powerful accuracy in using passively sensed data from mobile phones to predict everything from symptoms of Schizophrenia [67] to the likelihood of exercising [?] to incidents of binge drinking [9] to the likelihood of the screen of a smartphone being visible to its owner [18]. Collecting audio samples that are relevant to researchers and aligned with study participants' expectations requires sensing techniques that can detect anchor events accurately.

Across case studies, we also encountered moments when AAS provided an incomplete picture of participants' usage experience, highlighting the limits of collecting audio data alone. A number of existing systems have demonstrated the value of integrating multiple data streams (*e.g.*, [11, 40]), and future work to seamlessly stitch together audio data and system event logging would enhance the value of AAS. It would also be useful to combine audio recordings with screen capture or video data or to selectively capture audio by voice print.

6.3 Privacy, Participant Awareness, and Data Loss

In response to hesitation from participants about the intrusion of AAS, we iteratively evolved supports to enhance participant awareness of recordings. We found that when participants had the option to manually delete data themselves, 30% chose to do so one time over the course of four weeks, indicating both that this feature was valuable and that the lost data did not impact our ability to collect a large, rich dataset. Because participants' reflections suggested that they would make use of a more robust interface for reviewing and managing their data, we support participant-facing review of all data in our general-purpose AAS library. However, we have not yet evaluated the extent to which participants engage with this interface, the extent to which it leads to data loss, the extent to which it addresses participants' concerns, or the extent to which participants feel the need to have access to all data. Future design work remains to iterate on this UI and to study how diverse AAS studies can best meet users' privacy needs.

Future work also remains to understand the extent to which participants' behavior changes in response to real-time feedback notifying them of anchor events, and examining such changes in the context of a specific study could be a valuable standard practice in AAS. In Study 3 (where antecedent recording occurred before feedback was given), we did not see strong evidence that participants' behavior changed before and after a notification, though this was confounded by the fact that the anchor event was the wake word

"Alexa," which users spoke as they switched to a very different type of speech. And we did see hints that participants were aware, at least at times, of being recorded.

Prior work examining other *in situ* methods has shown that knowledge of data collection can create a Heisenberg Effect, such that the process of self-reporting experiences can change what participants do and how they feel about it [15]. Thus, by enhancing participants' awareness, the research team might increase participants' comfort and sense of control, but by diminishing participants' awareness of recordings, the research team might increase the ecological validity of the behaviors and responses they capture. Employing AAS necessitates wrestling with this tension in the context of the study in which it is employed.

Though the needs of different studies and study populations will vary, we argue for privileging participant awareness over participant authenticity, as it better honors the tenet of respect for persons [2]. In determining which form(s) of feedback to provide, researchers might consider how sensitive the audio data is likely to be, whether child participants will understand the recording process, whether an adult will be engaging with technology together with the child, and where participants are likely to be directing their attention during the anchor event. As Graue and Walsh [28] explain in discussing the ethics of research with children (p. 56):

> "*Entering other people's lives is intrusive. It requires permission, permission that goes beyond the kind that comes from consent forms. It is the permission that permeates any respectful relationship between people.*"

In deciding how forcefully to raise participants' awareness of recordings, how proactively to retrieve recordings after they have been collected locally, and how comprehensive to be in capturing recordings, researchers making use of AAS must carefully assess the specific permissions that truly have been granted to them by their participants.

7 CONCLUSION

Very young children are avid technology users who deserve to be first-class citizens in the design of systems and experiences that affect their lives. AAS is a data collection technique for field deployments that captures audio samples in response to anchor events of interest. This allows researchers to passively capture rich, qualitative feedback about moments of interest and the events that precede them, and to do so at scale with data that comes directly from children. By formalizing the principles of the approach and supporting these with an open-source library, we seek to empower the research community to confront the ethical tensions inherent in this work and to resolve them within the context of their own field deployments.

REFERENCES

[1] [n. d.]. Amazon Web Services (AWS) - Cloud Computing Services. https://aws.amazon.com/

[2] 1978. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research.* Superintendent of Documents.

[3] 2017. The Common Sense Census: Media Use by Kids Age Zero to Eight 2017 | Common Sense Media. , 64 pages. https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-kids-age-zero-to-eight-2017

[4] Saeed Abdullah, Elizabeth L. Murnane, Mark Matthews, Matthew Kay, Julie A. Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive rhythms. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*. ACM Press, New York, New York, USA, 178–189. https://doi.org/10.1145/2971648.2971712

[5] Sabirah Adams, Shazly Savahl, and Tobia Fattore. 2017. Children's representations of nature using photovoice and community mapping: Perspectives from South Africa. *International Journal of Qualitative Studies on Health and Well-being* 12, 1 (jan 2017), 1333900. https://doi.org/10.1080/17482631.2017.1333900

[6] William Albert, Thomas Tullis, and Donna Tedesco. 2009. *Beyond the usability lab: Conducting large-scale online user experience studies.* Morgan Kaufmann.

[7] AHMED S Alghamdi, ALIH Al-Badi, ROOBAEA ALROOBAEA, and P Mayhew. 2013. A Comparative Study of Synchronous and Asynchronous Remote Usability Testing Methods. *International Review of Basic and Applied Sciences* 1, 3 (2013), 61–97.

[8] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What happened to remote usability testing?. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, New York, New York, USA, 1405. https://doi.org/10.1145/1240624.1240838

[9] Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K. Dey, and Brian Suffoletto. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors* 83 (aug 2018), 42–47. https://doi.org/10.1016/J.ADDBEH.2017.11.039

[10] Erin Beneteau, Olivia Richards, Mingrui Zhang, Julie A. Kientz, Jason C. Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. 1–14. https://doi.org/10.1145/3290605.3300473

[11] Scott Carter, Jennifer Mankoff, and Jeffrey Heer. 2007. Momento. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, New York, New York, USA, 125. https://doi.org/10.1145/1240624.1240644 arXiv:arXiv:1303.5153v1

[12] Eve V Clark. 2009. *First language acquisition.* Cambridge University Press.

[13] Sunny Consolvo and Miriam Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (apr 2003), 24–31. https://doi.org/10.1109/MPRV.2003.1203750

[14] Juliet Corbin and Anselm Strauss. 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory. (2008).

[15] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* (2004), 175–182. https://doi.org/10.1145/985692.985715

[16] Pierre-Charles David and Thomas Ledoux. 2005. WildCAT. In *Proceedings of the 3rd international workshop on Middleware for pervasive and ad-hoc computing - MPAC '05*. ACM Press, New York, New York, USA,

1–7. https://doi.org/10.1145/1101480.1101483
[17] Anind K. Dey and Jennifer Mankoff. 2005. Designing mediation for context-aware applications. *ACM Transactions on Computer-Human Interaction* 12, 1 (mar 2005), 53–80. https://doi.org/10.1145/1057237.1057241
[18] Anind K. Dey, Katarzyna Wac, Denzil Ferreira, Kevin Tassini, Jin-Hyuk Hong, and Julian Ramos. 2011. Getting closer. In *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*. ACM Press, New York, New York, USA, 163. https://doi.org/10.1145/2030112.2030135
[19] Allison Druin. 1999. Cooperative inquiry. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*. ACM Press, New York, New York, USA, 592–599. https://doi.org/10.1145/302979.303166
[20] Allison Druin. 2002. The role of children in the design of new technology. *Behaviour & Information Technology* 21, 1 (jan 2002), 1–25. https://doi.org/10.1080/01449290110108659
[21] Nicolas Ducheneaut and Robert J. Moore. 2004. The social side of gaming. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*. ACM Press, New York, New York, USA, 360. https://doi.org/10.1145/1031607.1031667
[22] Jóhanna Einarsdóttir. 2007. Research with children: methodological and ethical challenges. *European Early Childhood Education Research Journal* 15, 2 (jun 2007), 197–211. https://doi.org/10.1080/13502930701321477
[23] Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa Al'Absi, and Siddharth Shah. 2011. AutoSense. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems - SenSys '11*. ACM Press, New York, New York, USA, 274. https://doi.org/10.1145/2070942.2070970
[24] Allison Farber, Allison Druin, Gene Chipman, Dawn Julian, and Sheila Somashekher. 2002. How Young Can Our Technology Design Partners Be?. In *PDC*. 272–277.
[25] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience. In *Proceedings of the 5th international conference on Mobile systems, applications and services - MobiSys '07*. ACM Press, New York, New York, USA, 57. https://doi.org/10.1145/1247660.1247670
[26] Amy Gallo. 2017. A Refresher on A/B Testing. *Harvard Business Review* (jun 2017). https://hbr.org/2017/06/a-refresher-on-ab-testing
[27] Franca Garzotto. 2008. Broadening children's involvement as design partners. In *Proceedings of the 7th international conference on Interaction design and children - IDC '08*. ACM Press, New York, New York, USA, 186. https://doi.org/10.1145/1463689.1463755
[28] M Elizabeth Graue and Daniel J Walsh. 1998. *Studying children in context: Theories, methods, and ethics*. Sage Publications.
[29] Deborah Harcourt. 2011. An encounter with children: Seeking meaning and understanding about childhood. *European Early Childhood Education Research Journal* 19, 3 (sep 2011), 331–343. https://doi.org/10.1080/1350293X.2011.597965
[30] Gillian R. Hayes, Shwetak N. Patel, Khai N. Truong, Giovanni Iachello, Julie A. Kientz, Rob Farmer, and Gregory D. Abowd. 2004. The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid. Springer, Berlin, Heidelberg, 168–179. https://doi.org/10.1007/978-3-540-28637-0_15
[31] Gillian R. Hayes, Khai N. Truong, Gregory D. Abowd, and Trevor Pering. 2005. Experience buffers. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*. ACM Press, New York, New York, USA, 1435. https://doi.org/10.1145/1056808.1056935
[32] Morten Hertzum, Pia Borlund, and Kristina B. Kristoffersen. 2015. What Do Thinking-Aloud Participants Say? A Comparison of Moderated and Unmoderated Usability Sessions. *International Journal of Human-Computer Interaction* 31, 9 (sep 2015), 557–570. https://doi.org/10.1080/10447318.2015.1065691
[33] Alexis Hiniker, Sharon S. Heung, Sungsoo (Ray) Hong, and Julie A. Kientz. 2018. Coco's Videos. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, Vol. 2018-April. 1–13. https://doi.org/10.1145/3173574.3173828
[34] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A Retrospective Memory Aid. Springer, Berlin, Heidelberg, 177–193. https://doi.org/10.1007/11853565_11 arXiv:9780201398298
[35] Karen Holtzblatt and Sandra Jones. 1993. Contextual inquiry: A participatory technique for system design. *Participatory design: Principles and practices* (1993), 177–210.
[36] John Hughes, L Bannon, J Bowers, P Carstensen, K Kuutti, J Pycock, T Rodden, K Schmidt, D Shapiro, W Sharrock, and Others. 1993. Informing CSCW system requirements. *COMIC, Esprit Basic Research Project* 6225 (1993).
[37] Stephen S. Intille, John Rondoni, Charles Kukla, Isabel Ancona, and Ling Bao. 2003. A context-aware experience sampling tool. In *CHI '03 extended abstracts on Human factors in computer systems - CHI '03*. ACM Press, New York, New York, USA, 972. https://doi.org/10.1145/766098.766101
[38] Sara Isola and Jerry Fails. 2012. Family and Design in the IDC and CHI Communitieis. *IDC 2012* (2012), 40–49.
[39] Janne J. Jensen and Mikael B. Skov. 2005. A review of research methods in children's technology design. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05*. ACM Press, New York, New York, USA, 80–87. https://doi.org/10.1145/1109540.1109551
[40] Jasmine Jones, David Merritt, and Mark S. Ackerman. 2017. KidKeeper. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. ACM Press, New York, New York, USA, 1864–1875. https://doi.org/10.1145/2998181.2998348
[41] Hilda K Kabali, Matilde M Irigoyen, Rosemary Nunez-Davis, Jennifer G Budacki, Sweta H Mohanty, Kristin P Leister, and Robert L Bonner. 2015. Exposure and Use of Mobile Media Devices by Young Children. *Pediatrics* (nov 2015), peds.2015–2151–. https://doi.org/10.1542/peds.2015-2151
[42] Eiman Kanjo. 2010. NoiseSPY: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications* 15, 4 (aug 2010), 562–574. https://doi.org/10.1007/s11036-009-0217-y
[43] Kostadin Kushlev and Elizabeth W. Dunn. 2018. Smartphones distract parents from cultivating feelings of connection when spending time with their children. *Journal of Social and Personal Relationships* (apr 2018), 026540751876938. https://doi.org/10.1177/0265407518769387
[44] Reed Larson and Mihaly Csikszentmihalyi. 1983. The experience sampling method. *New directions for methodology of social & behavioral science* (1983).
[45] Kai Lukoff, Cissy Yu, Julie Kientz, and Alexis Hiniker. 2018. What Makes Smartphone Use Meaningful or Meaningless? *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (mar 2018), 1–26. https://doi.org/10.1145/3191754
[46] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys that Listen. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17 (CHI '17)*. ACM, New York, 5197–5207. https://doi.org/10.1145/3025453.3025735
[47] David R. Millen. 2000. Rapid ethnography. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, and techniques - DIS '00*. ACM Press, New York, New York, USA, 280–286.

https://doi.org/10.1145/347642.347763
[48] E Olmsted and M Gill. 2005. In-person usability study compared with self-administered web (remote-different time-place) study: does mode of study produce similar results. In *Proceedings of UPA*.
[49] Judith S Olson and Wendy A Kellogg. 2014. *Ways of Knowing in HCI*. Springer Science & Business.
[50] Antti Oulasvirta. 2009. Field Experiments in HCI: Promises and Challenges. In *Future Interaction Design II*. Springer London, London, 87–116. https://doi.org/10.1007/978-1-84800-385-9_5
[51] Leysia Palen and Marilyn Salzman. 2002. Voice-mail diary studies for naturalistic data capture under mobile conditions. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work - CSCW '02*. ACM Press, New York, New York, USA, 87. https://doi.org/10.1145/587078.587092
[52] Kiran K. Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J. Rentfrow. 2011. SociableSense. In *Proceedings of the 17th annual international conference on Mobile computing and networking - MobiCom '11*. ACM Press, New York, New York, USA, 73. https://doi.org/10.1145/2030613.2030623
[53] Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J. Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense. In *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*. ACM Press, New York, New York, USA, 281. https://doi.org/10.1145/1864349.1864393
[54] Mika Raento, Antti Oulasvirta, and Nathan Eagle. 2009. Smartphones: An emerging tool for social scientists. *Sociological Methods and Research* 37, 3 (feb 2009), 426–454. https://doi.org/10.1177/0049124108330005 arXiv:arXiv:1011.1669v3
[55] M. Raento, A. Oulasvirta, R. Petit, and H. Toivonen. 2005. ContextPhone: A Prototyping Platform for Context-Aware Mobile Applications. *IEEE Pervasive Computing* 4, 2 (apr 2005), 51–59. https://doi.org/10.1109/MPRV.2005.29
[56] Robert W. Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An Experience Sampling Study of User Reactions to Browser Warnings in the Field. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–13. https://doi.org/10.1145/3173574.3174086
[57] Katharina Reinecke and Abraham Bernstein. 2011. Improving performance, perceived usability, and aesthetics with culturally adaptive user interfaces. *ACM Transactions on Computer-Human Interaction* 18, 2 (jun 2011), 1–29. https://doi.org/10.1145/1970378.1970382
[58] Yvonne Rogers, Kay Connelly, Lenore Tedesco, William Hazlewood, Andrew Kurtz, Robert E. Hall, Josh Hursey, and Tammy Toscos. 2007. Why It's Worth the Hassle: The Value of In-Situ Studies When Designing Ubicomp. In *UbiComp 2007: Ubiquitous Computing*. Springer Berlin Heidelberg, Berlin, Heidelberg, 336–353. https://doi.org/10.1007/978-3-540-74853-3_20
[59] Juergen Sauer, Andreas Sonderegger, Klaus Heyden, Jasmin Biller, Julia Klotz, and Andreas Uebelbacher. 2019. Extra-laboratorial usability tests: An empirical comparison of remote and classical field testing with lab testing. *Applied Ergonomics* 74 (jan 2019), 85–96. https://doi.org/10.1016/j.apergo.2018.08.011
[60] Abigail J. Sellen, Andrew Fogg, Mike Aitken, Steve Hodges, Carsten Rother, and Ken Wood. 2007. Do life-logging technologies support memory for the past?. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, New York, New York, USA, 81. https://doi.org/10.1145/1240624.1240636
[61] Clayton Shepard, Ahmad Rahmati, Chad Tossell, Lin Zhong, and Phillip Kortum. 2011. LiveLab. *ACM SIGMETRICS Performance Evaluation Review* 38, 3 (jan 2011), 15. https://doi.org/10.1145/1925019.1925023
[62] Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. 2008. A diary study of mobile information needs. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, New York, New York, USA, 433. https://doi.org/10.1145/1357054.1357125
[63] Arthur A Stone, Saul S Shiffman, and Marten W DeVries. 1999. Ecological momentary assessment. (1999).
[64] Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
[65] Caroline Wang and Mary Ann Burris. 1997. Photovoice: Concept, Methodology, and Use for Participatory Needs Assessment. *Health Education & Behavior* 24, 3 (jun 1997), 369–387. https://doi.org/10.1177/109019819702400309
[66] Rui Wang, Emily A. Scherer, Vincent W. S. Tseng, Dror Ben-Zeev, Min S. H. Aung, Saeed Abdullah, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, and Michael Merrill. 2016. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*. ACM Press, New York, New York, USA, 886–897. https://doi.org/10.1145/2971648.2971740
[67] Rui Wang, Emily A. Scherer, Megan Walsh, Weichen Wang, Min Hane Aung, Dror Ben-Zeev, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, and John Kane. 2018. PREDICTING SYMPTOM TRAJECTORIES OF SCHIZOPHRENIA USING MOBILE SENSING. *GetMobile: Mobile Computing and Communications* 22, 2 (sep 2018), 32–37. https://doi.org/10.1145/3276145.3276157
[68] Graver J. Whitehurst and Christopher J. Lonigan. 1998. Child Development and Emergent Literacy. *Child Development* 69, 3 (jun 1998), 848–872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x
[69] Jason C. Yip, Kiley Sobel, Caroline Pitt, Kung Jin Lee, Sijin Chen, Kari Nasu, and Laura R. Pina. 2017. Examining Adult-Child Interactions in Intergenerational Participatory Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, New York, New York, USA, 5742–5754. https://doi.org/10.1145/3025453.3025787