Why Did you Stop? - Investigating Origins and Effects of Interruptions during Mobile Language Learning

Fiona Draxler*
Christina Schneegass*
fiona.draxler@ifi.lmu.de
christina.schneegass@ifi.lmu.de
LMU Munich
Munich, Germany

Jonas Safranek
LMU Munich
Munich, Germany
safranek.jonas@gmail.com

Heinrich Hussmann
LMU Munich
Munich, Germany
hussmann@ifi.lmu.de

ABSTRACT

The technological advances of smartphones facilitate the transformation of learning from the classroom to an activity that can happen anywhere and anytime. While micro-learning fosters ubiquitous learning, this flexibility comes at the cost of having an uncontrolled learning environment. To this point, we know little about the usage of mobile learning applications, particularly the occurrence of interruptions and the harm they cause. By diverting users' attention away from the learning task, interruptions can potentially compromise learning performance. We present a four-week in-the-wild study ($N = 12$) where we investigate learning behavior and the occurrence of interruptions based on device logging and experience sampling questionnaires. We recorded 276 interruptions in 327 learning sessions and found that interruption type as well as users' context influence learning sessions and the severity of the interruption (i.e., session termination likeliness). We discuss challenges and opportunities for the design of automated mechanisms to detect and mitigate interruptions in mobile learning.

CCS CONCEPTS

• **Human-centered computing** → *Field studies*; *Mobile devices*; **Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

Mobile Learning, Interruptions, Task Resumption Support, Experience Sampling Method, Empirical Study

ACM Reference Format:
Fiona Draxler, Christina Schneegass, Jonas Safranek, and Heinrich Hussmann. 2021. *Why Did you Stop?* - Investigating Origins and Effects of Interruptions during Mobile Language Learning. In *Mensch und Computer 2021 (MuC '21), September 5–8, 2021, Ingolstadt, Germany*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3473856.3473881

*Both authors contributed equally to this research.

MuC '21, September 5–8, 2021, Ingolstadt, Germany

ACM ISBN 978-1-4503-8645-6/21/09…$15.00
https://doi.org/10.1145/3473856.3473881

1 INTRODUCTION

The ubiquity of mobile devices has opened up new possibilities for learning practice that is integrated into our everyday lives instead of limited to classrooms. Learning programming while waiting for the bus or learning a new language while sitting in the doctor's waiting room have become common scenarios. Technological advances enable the design of new learning practices (e.g., micro-learning) but also lead to considerable challenges [26]. Mobile learning (ML) is convenient because it is independent of time and location [8]. However, the stimuli present in uncontrolled environments can easily distract learners. Even the devices that learners use (i.e., smartphones or tablets) add an additional source of interruptions. They are built to foster multitasking and present an increasing number of (distracting) notifications over the course of the day [38]. Those continuous distractions, caused by either the environment or our devices, can interrupt and harm the learning experience. Research has shown that interruptions can affect task performance, leading to higher error rates or impaired memory consolidation [3, 9, 22, 25, 29]. While some studies investigated how people use ML applications [8, 42], we still know very little about interruptions that occur and how (avoidable) interruptions could be mitigated to support the learners' continued focus [15].

This work aims to provide additional, quantified evidence of everyday ML app usage and, in particular, to evaluate what interruptions actually occur as well as how this influences learning behavior. Therefore, we developed a customized Android app called "Learning Activity and Interruption Recognition Application" (LAIRA). This app logs learning behavior, the users' context, and potentially interrupting smartphone actions and events.

We deployed the app in a four-week study field ($N = 12$), where we recorded 327 learning sessions. Of these, participants supplemented 266 with post-hoc reports through experience sampling questionnaires (ESQs). Our results show that users frequently encounter interruptions in ML situations (276 interruptions in 327 learning sessions). Interruptions are caused by environmental stimuli, by the users themselves, or by the mobile device (mainly originating from messaging apps). Our participants reported that external interruptions had the highest potential for distraction, leading to the highest reported number of terminations of learning sessions (37 out of 99).

While the participants reported that 80% of their interruptions were not of high importance and could be postponed or ignored, many of them still led to the termination of the learning session (around 37%). Especially the frequent occurrence of detectable device interruptions and the high percentage of low-priority interruptions suggests that there is an opportunity to support mobile learners in reducing or mitigating interruptions.

By obtaining information on the origins and effects of interruptions, we aim to create a basis for developing automated mechanisms to predict and strategies to mitigate such interruptions and allow for more focused and productive ML sessions. Hereby, ML could rise from teaching micro-contents in micro-interactions to the more complex learning content, such as STEM topics.

Contribution statement: The contribution of our work is threefold: We (1) present an in-depth view on data of language learning app users from an field study and (2) investigate the occurrences and effects of interruptions on users and their learning sessions. We further (3) developed the LAIRA application and implemented a mechanism for identifying different types of interruptions.

2 BACKGROUND AND RELATED WORK

2.1 Mobile Learning in Context

Mobile devices enable us to learn anywhere and at any time [8]. By using a micro-learning approach, ML apps can deliver small units of learning content so that users can learn on the go [8, 14]. Indeed, usage of ML apps is common in a number of different scenarios [13, 42]. For example, learning can happen in idle moments such as waiting situations [14], but also on commutes, at home, at work, or in public spaces [8, 42]. Analyses of ML app usage habits further show that learning sessions of mobile language learners typically range between 5 and 20 minutes [42]. The variety of learning contexts at different locations entail context characteristics such as varying levels of noise in the environment, the learners' company, stress levels, privacy requirements, and planning of learning sessions [42]. Ideally, ML apps should be designed such that they can promote learning in any possible environment. In order to assess learner needs in different contexts, detailed analysis of context factors and their effect on learning is necessary. Currently available analyses tend to only be partial summaries of available data (e.g., [6]), or to be based on interviews and surveys (e.g., [8, 42]), while more comprehensive log data on ML app usage is scarce.

2.2 Interruptions in Mobile Learning

Inevitably, the diversity of learning contexts goes hand in hand with a diversity of interruptions that can occur during the learning session. Prior research on interruptions shows that disrupting a task can affect its error rate and completion time [3, 22, 25]. In particular, studies show that longer interruptions can have more severe effects, as users require more time to resume their original task [18, 30]. Moreover, interruptions can have a strong negative impact especially for complex tasks that require the user to keep more information in their working memory, as the task state needs to be recovered after an interruption [21]. Thus, ML apps are typically limited to content types that do not require long learning streaks but benefit from high repetition counts (e.g., vocabulary in language learning [11]), as they are rather robust with respect to interruptions. Designing ML apps for learning content that is more complex and requires continued attention remains challenging. Generally speaking, any situation in which an action or event ("secondary task") shifts the learners' focus of attention away from the learning task ("primary task") can be considered an interruption. Such interruptions can be caused by the devices users learn with, i.e., the smartphone or tablet (*device-internal interruptions*), by external stimuli such as noises or people approaching (*external interruptions*), or can be rooted internally, for example when a user gets tired, decides to check social media, or when the mind starts wandering (*internal interruptions*) [22, 41].

2.3 Avoiding Interruptions and Mitigating their Effects

To avoid negative effects on the learning, interruptions in ML should be avoided whenever possible. One solution that has been proposed in this context is the implementation of notification managers that defer device interruptions by delaying push notifications to activity breakpoints [16, 20, 31, 34]. However, internal and external interruptions are difficult to predict with on-device sensing. In this case, supporting task resumption can help attenuate the negative effects of interruptions. For example, Oulasvirta & Saariluoma (2006) highlight that interface organization can facilitate the encoding of the workspace in users' memory and hereby help them to quickly find their way back to the original task [32, 33]. In addition, simple highlights (e.g., [7, 24, 27, 28, 46]) or more explicit memory cues that restore even complex task context (e.g., [39, 44, 49]) can be used. Even in everyday applications we can find features that help us resume a task at a later point in time. Overviews of recently opened documents, reminder emails, or opening a document at the last point of editing can positively influence the resumption of a task. In the specific use case of ML, suggested techniques for task resumption support after an interruption include focus exercises, reminders, summaries, or memory/mnemonic cues [15]. Further, prior work has compiled suggestions for the design of task resumption cues in ML-based on the analysis of literature from related domains. The authors emphasize the importance of considering the variety of usage situations for the design and to make use of the interruption lag if possible [41]. The latter refers to utilizing the time between the interruption and the resumption of the original task, namely, by applying task resumption cues to guide the user.

2.4 Research Gap

While prior work has thoroughly investigated ML applications from a user-reported perspective, quantifiable data on the interaction with ML apps is still sparse. Further, reasons for interruptions as well as their (potentially negative) effects on the learning performance have yet to be explored in more detail. Although interruptions and task resumption cues have been well-researched in stationary settings, their effectiveness depends on their adaptation to the device capabilities, the task at hand, and the interruptions that occur. Thus, existing task resumption support techniques cannot easily be generalized and applied to an ML scenario. With this work, we aim to investigate ML interruptions in the wild in order to derive suitable mitigation strategies.

3 CONCEPT AND IMPLEMENTATION OF LAIRA

We developed a custom "Learning Activity and Interruption Recognition Application" (LAIRA) to gather data on users' interaction with learning apps, their current context, and interruptions that occur during learning. It logs learning sessions and various device events and augments the recorded information with an experience sampling questionnaire (ESQ) issued after each learning session.

LAIRA was developed for Android 7 or higher. Context data are collected from device sensors and interaction patterns; some additional data are retrieved using the open-source AWARE Framework[1]. For LAIRA to log triggered push notifications and package names of all opened apps including widgets and launchers, users must grant Accessibility permissions. All session data are stored in a Google Firestore Database[2]. Following the recommendations in [10], we implemented a combination of *BroadcastReceivers* and *JobServices* to keep the app running at all times despite various battery optimization techniques applied by different device vendors.

3.1 App Interface

Interaction with LAIRA is limited to a study dashboard, a timeline view of recorded sessions, and an integrated survey view (see Figure 1). In the study dashboard, users select a learning app from a drop-down list of all apps on their phones. This is necessary for tracking learning sessions (see Section 3.2 for more details). In addition, the dashboard gives an overview of permissions granted to LAIRA (colored green in the figure). The timeline view is intended to increase the trust of users by showing them what data are recorded, as the required permissions could potentially be used for privacy-invading purposes. Finally, the survey view shows an initial and final survey at the beginning and end of the study period, respectively. Thanks to this coupling, the survey results and app interaction data are connected without compromising the anonymity of participants.

3.2 Event Logging

Event logging starts once users have selected a learning app in the study dashboard. Every learning session LAIRA recognizes is saved to the database, including relevant general meta-data such as the current app name, a time stamp, learning session or event length, learning session ID, and user ID. The *learning session* events include the following information:

- **Learning App** - Indicating the name of the learning application in use.
- **Session Duration** - Measuring the time from the start to the end of a learning session. We define the end of a learning session either as an active closing action by the user (learning app moved to the background or screen turned off) or when the time threshold of inactivity is reached. This time frame in LAIRA is ten minutes and reflects the short learning sessions common in ML as well as the decay of goals and problem states in working memory. When the user returns to the learning activity after long periods of inactivity, we consider this as the start of a new learning session.

We gather additional information about events that can cause *interruptions* or cause learners to terminate their learning sessions. If multiple interrupting events occur during a single learning session, each interruption is registered as an individual event. Only the last interruption will be inquired upon in the ESQ to keep the effort for the user to a minimum.

- **Notifications & Communication** - This includes push notifications, SMS, and phone calls. We do not process or store any text or voice content but only application package names and metadata. To catch SMS and phone calls, we register a *BroadcastReceiver* and set actions for *android.intent.action.PHONE_STATE* (in particular *RINGING*) and *android.provider.Telephony.SMS_RECEIVED*. To collect data on incoming push notifications, we implement a *NotificationListenerService*[3] and store the app name, the notification priority, and check if it caused a sound or vibration. The priority (as well as the sound and vibration on Android 7.1 and lower) can influence whether or not the notification is displayed as a heads-up notification[4], which is more likely to distract users than less intrusive notification types.
- **Application Switches** - Switches of applications on the phone can occur for different reasons. If the user switches from the learning app to a different app without prior indication (i.e., a notification), we label the switch as an internal interruption. In this case, we assume that the user decided to start another activity on their own accord. For example, users might want to quickly put an item on a digital shopping list. If the user switches apps due to an SMS, call, or notification, we consider the interruption event triggered by the device.
- **Screen Locks** - Every time the user actively locks their screen or the screen is locked by the phone moving to an idle state, a screen lock event is recorded. The screen lock could indicate an external interruption that cannot be tracked or that the user ended the learning session. We record these events as ambiguous interruptions. Their cause has to be confirmed by the user in the ESQ, otherwise, the label "ambiguous" remains.

All of the events listed above can cause the LAIRA app to register a "Session_End" event. Based on the flow depicted in Figure 2, the app categorizes the interruption types and triggers the ESQ. Further, the following *context information* is acquired:

- **Movement type** - We record movement types obtained from the Google Activity Recognition API [5], i.e., *IN_VEHICLE*, ON_BICYCLE, ON_FOOT, RUNNING, *WALKING*, *STILL*, or *UNKNOWN* . To assure sufficient data quality, we only include movements with a confidence level of 90 or greater. A

[1]https://awareframework.com/, last accessed April 15th, 2021
[2]https://firebase.google.com/docs/firestore, last accessed April 15th, 2021
[3]https://developer.android.com/reference/android/service/notification/NotificationListenerService, last accessed April 15th, 2021
[4]A floating window that is shown at the top of the screen for a short moment when the device is unlocked, https://developer.android.com/guide/topics/ui/notifiers/notifications, last accessed April 15th, 2021
[5]https://developers.google.com/location-context/activity-recognition, last accessed April 15th, 2021

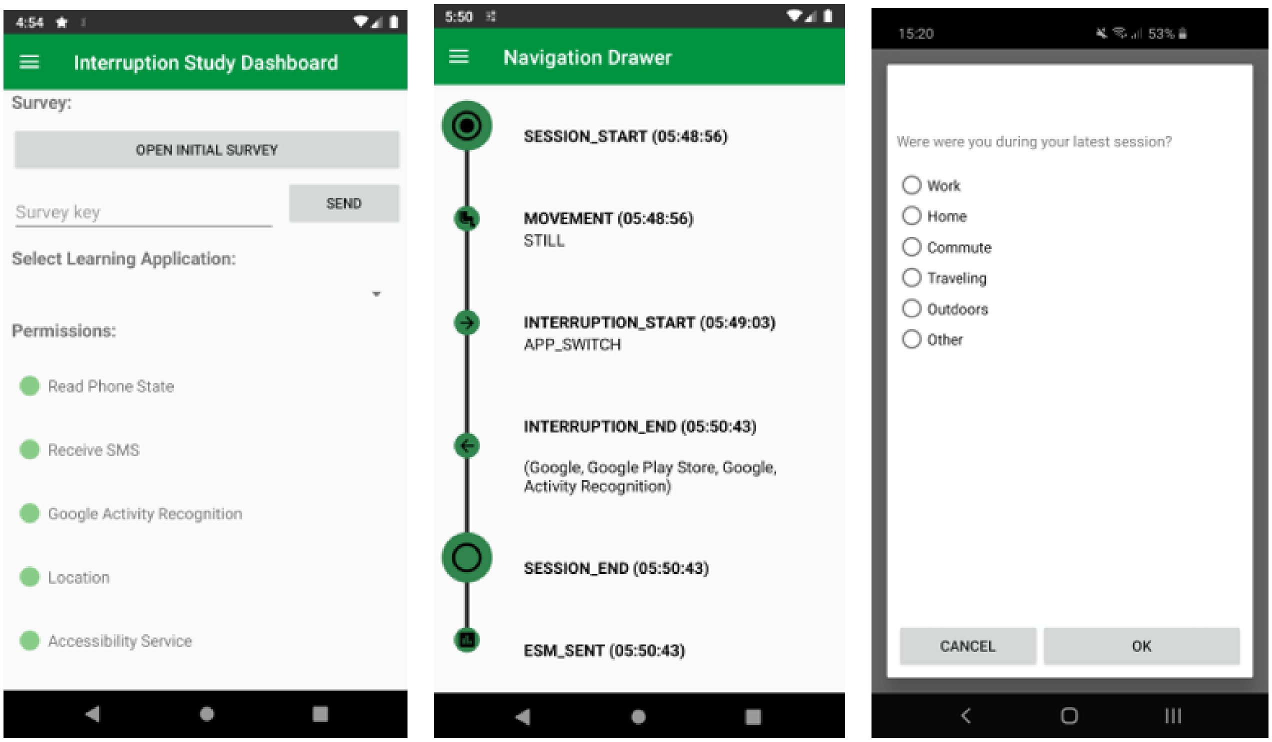


Figure 1: Left: The main view of the LAIRA application showing the link to the initial survey, the survey user key input field, the drop-down list to select a learning application to track, and the permissions users have granted to the app. Middle: The timeline view of all recorded user data. Right: An exemplary excerpt of the ESQ prompted after each learning session.

movement type is considered active until a movement type switch occurs.

3.3 Experience Sampling

We further collected self-reported data on learning context and interruptions through the experience sampling (ES) method [4]. To create the ES prompts, we used the *ESMFactory* class provided with the *AWARE* framework. LAIRA triggers new experience sampling questionnaires (ESQs) via a push notification ten minutes after the last interaction with a learning app was recorded. The ESQs only comprise multiple-choice questions that allow for quick completion; some of the questions are dynamically adapted based on recorded events. Clicking on the ES notification opens a pop-up dialog with the following questions on the learning session. The questions used for determining the interruption type are also displayed in Figure 2.

- **Where were you during your latest session?**
  Options: Work | Home | Commute | Travelling | Outdoors | Other
- **Were you alone or in company during your latest session?**
  Options: Alone | With one other person | With more than one other person
- **Please confirm the movement type we detected.** (see Google Activity Recognition API)
- If the user received at least one notification during the recorded learning session, we further ask for confirmation of distraction:
  Did you receive any distracting notifications during your latest session?
  Options: Yes | No
- **Why did you end your learning session?**
  Options: *Device Internal* – "I was interrupted by something on my phone (e.g., a notification, call, SMS, email, etc.)" | *External* - "I was distracted by something external to myself or the phone (e.g., doorbell, other people, having to get off of train, etc.)" | *Internal* - "I was distracted internally (e.g., tiredness, could not concentrate, thinking of something else, mind-wandering, etc.)" | *Intentional* – "I was done using the app."
- If the user did not answer with "intentional", a follow-up question is shown: **How important was it that you follow up upon the interruption?**
  Options: "Very important- it was urgent / time-critical" | "Moderate - I had to do it eventually in the near future" | "Not important - I could have ignored it and continued learning"

In the AWARE ESMFactory, users can always dismiss ESQ notifications and they do not have to answer right away. If the user does not fill in the ESQ within three hours after a learning session, it is discarded to avoid bias due to fading memory of the learning session (cf. [5]). We chose this time window in line with findings from prior

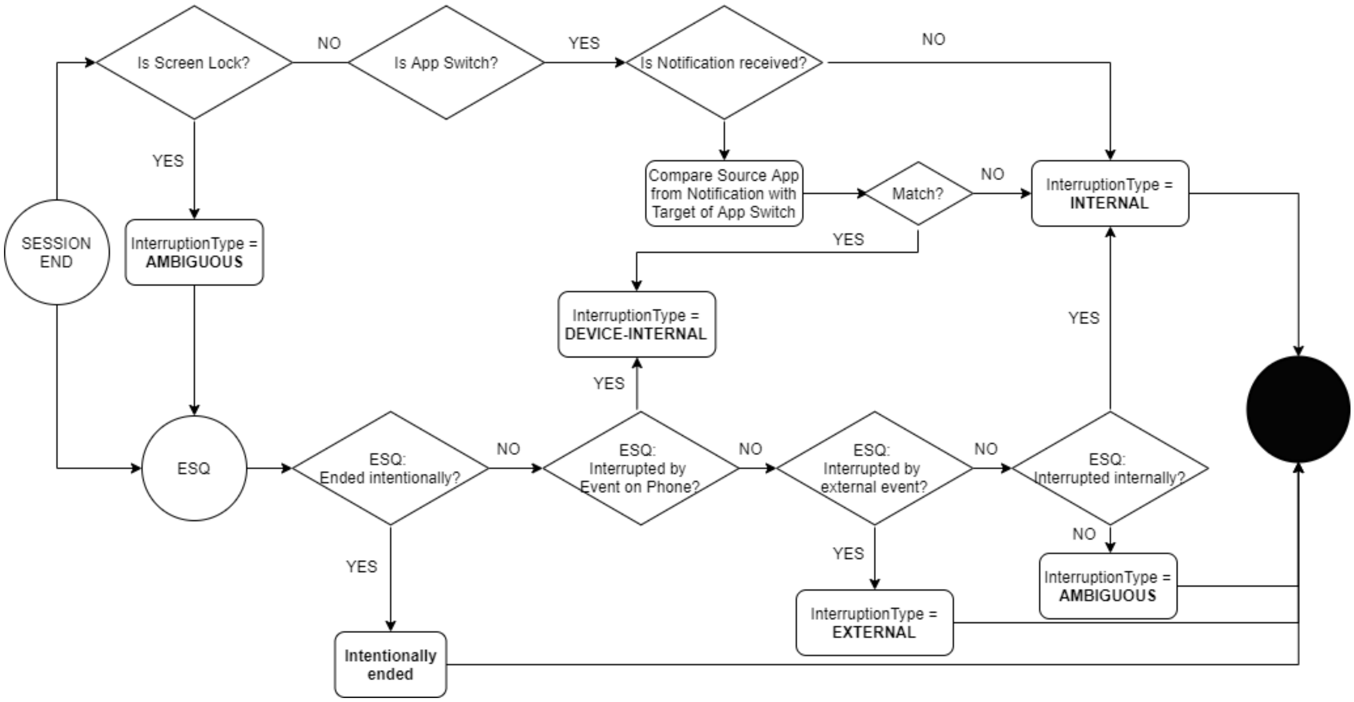


Figure 2: This flow chart depicts the LAIRA app's process of categorizing interruptions that terminated the learning session once the Session_End event has been registered and the ESQ answered, dismissed, or expired.

work indicating that mobile learning apps are rarely used multiple times per day but rather daily or less than daily [42]. Therefore, we aim for high ESQ participation and expect participants to remember the context of their learning session within a three-hour window. Additionally, as van Berkel et al. recommend, a previously unanswered ESQ is deleted if the user finishes a new learning session before answering the ESQ of the previous session [5].

4 FIELD STUDY ON MOBILE LEARNING AND INTERRUPTIONS

4.1 Study Design

We assessed the learning habits of ML app users in regards to the timing and duration of learning sessions, as well as interruptions during learning in a four-week field study using the LAIRA app. To this end, we collected quantitative ML app usage data as well as characteristics of the respective learning context (e.g., movement type and received notifications). The quantitative data were augmented with responses to experience sampling questionnaires (ESQs). We evaluated occurring interruptions caused by the *device* (e.g., notifications, calls), *external* circumstances (e.g., noise, distractions), or *internal* reasons (e.g., getting tired, mind-wandering, pausing to take care of a forgotten to-do). The first part of our analysis describes the characteristics of ML sessions and interruptions during learning. It is guided by the following research questions:

- *RQ* 1 How do people use mobile language learning apps?
- *RQ* 2 How often and in what contexts do interruptions of the three types (internal, external, device) occur during ML?

Further, we investigate potential influences of the interruption type, length, and the contexts of users (see RQ 3-5) on the learning application usage. Thus, we aim to derive implications for interruption mitigation.

- *RQ* 3 Does the interruption type influence the length of interruptions and the risk of terminating a learning session?
- *RQ* 4 Do interruption length and count impact the total net length of learning sessions?
- *RQ* 5 Does the context of participants influence the duration of learning sessions and the number or length of interruptions?

4.2 Procedure

After registering for our study, participants received an email with detailed instructions for the installation of the LAIRA app and the study procedure. Furthermore, we provided participants with information on our university's data protection regulation and asked them to read and sign the consent form. To allow for remote execution of our study, upon giving informed consent, users could start the initial questionnaire from within the application. Immediately after the successful installation of the app, the user was prompted with the link to the first questionnaire asking for demographic information and previous experience with ML applications. The questionnaire responses and app logging data were linked using an individual, automatically generated user token.

The participants were encouraged to use their ML application of choice and follow their usual learning routine. After each learning session, we asked them to fill in a brief ES questionnaire to gather additional information on the learning session, in particular, the reasons for ending the sessions and potential interruptions (see Section 3.3 for details). At the end of the four-week study, the application informed the participants that they had reached the end of the study and prompted them with the link to the final survey from within the application.

The LAIRA app does not process notification content or caller names; it only stores the respective app package name. Nonetheless, due to the processing of semi-personal data (i.e., movement types, apps used), we acquired approval from our university's ethics committee[6] to perform this study.

For their successful participation, all people received a 25 Euro voucher for an online store or an equivalent amount of study credit points.

4.3 Sample

We recruited 12 participants using our university's mailing lists and social media channels. One participant did not hand in the final questionnaire but used the app correctly over the full course of the study. We include this participant's data in the description of the logging but report the questionnaire results only for the subset of eleven.

The participants' age ranged from 19 to 60 (M = 27.84, SD = 10.43) years and they all identified as female. Four stated that their highest degree was a high school diploma, three had a bachelor's degree, four a master's degree, one person reported a lower-than-high–school degree. Five participants were full-time students, one was working full-time, four were studying and working part-time, and two were unemployed at the time of the study[7].

To assess the participants' smartphone usage behavior, we asked them to specify their average smartphone, social media, and messaging habits (if they were unsure, we advised them to use their phone's digital well-being feature showing an overview of screen time by app category). In particular, they could select time ranges from 0-15, 15-30, 31-60, 61-120, 121-180, and 180+ minutes per day. For general smartphone usage, the majority of participants (11) selected phone usage times between 60 and 181+ minutes per day. When reporting on social media usage in particular, the answers ranged from 0-15 (2) to 61-120 min (4), the majority (6) being in between. For messaging applications, most of the participants selected the ranges 16-30 min (5) and 31-60 min (4). The remaining three participants reported up to 120 min daily usage. The participants also stated that they received a mean of 133.58 push notifications per day (SD = 127.50, estimated or looked up via the digital well-being feature), with a maximum of 400 and a minimum of three.

All our study participants reported having prior experience with ML applications. Two participants were currently using the apps extensively, five currently but only occasionally. The remaining participants had used ML apps in the past, four extensively and one rather sporadically.

[6]Ethical approval granted: https://www.mathematik-informatik-statistik.uni-muenchen.de/ethikkommission/index.html; case number EK-MIS-2020-019

[7]As this study ran during the COVID-19 pandemic, we further asked the participants to specify their working situation. All participants who were currently working or studying stated to be able to do so from home. We will discuss the implications of the situation on our study results in the Limitation Section.

4.4 Results

In this section, we present the results of our user study grouped by the research questions introduced in Section 4.1. For RQ1 and RQ2 we present descriptive statistics; for RQ3 and RQ4, we apply hypothesis tests and calculate correlations. Finally, for RQ5, we fit regression models. There were large differences between the learning habits of individual participants and the number or learning sessions per participant. Therefore, for hypothesis testing, we report Bayesian ANOVAs and post-hoc tests where we control for the individual participants' as random effects[8]. The Bayesian tests additionally allow us to draw statistical conclusions even on small sample sizes (cf. [23]). These measures were computed with JASP [48]. In the section on RQ5, we apply Generalized Additive Models for Location Scale and Shape (GAMLSS) that predict our response variables based on context data. This approach is similar to linear mixed models but allows for modeling based on skewed distributions [45]. Again, we integrate the participants as random effect. The models were computed with R [35] and the GAMLSS package [36].

4.4.1 Characteristics of Learning Sessions (RQ1). In total, we recorded 328 learning sessions with LAIRA, the majority using the language learning apps Duolingo (218 learning sessions), Babbel (39), Drops (39), or Memrise (10). An additional 22 sessions were recorded on the learning app Quizlet, with which the user can design flashcards for any learning topic. All apps were rated as enjoyable (M = 4.09, SD = 0.51 on a 5-point Likert scale with 1 as the worst and 5 as the best possible value) and participants reported a good user experience with their learning apps (M = 4.09, SD = 0.79) and the ESQ application (M = 3.82, SD = 0.57). Overall, they were satisfied with their personal learning progress over the course of this study (M = 3.73, SD = 0.75). Therefore, we do not expect that the choice of ML app and the individual learning behavior affected the validity of the collected data.

The duration of learning sessions ranged from 27 seconds to 3223 seconds with one outlier of 11278 seconds (approximately 3h). This may have been caused by a deactivation of the automated screen locking mechanism. Hence, we excluded this session from all further analyses. The resulting median session length is 487.6s (8:07.6 minutes, M = 671.4s, SD = 577.0s). Participants completed between eight and 46 learning sessions (M = 27.3, SD = 11.0). Figure 4 shows that the number of learning sessions slightly decreased towards the end of the 4-week period. Four additional learning sessions of three different participants were recorded after the official duration of 28 days because logging only stopped once the final survey was completed. We did not exclude these sessions, as the specific days were not relevant for our analysis.

Participants supplemented 266 of the remaining 327 learning sessions with additional data through the ES questionnaires (ESQs). In the remaining 61 cases, the ESQs were either dismissed or removed after not being completed in the three-hour time window after the learning session. It has to be noted that not all ESQs were

[8]Even for tests with two conditions only, we used Bayesian ANOVAs instead of t-tests in order to control for random effects. The reported Bayes factors BF_{10} indicate the likelihood ratio of the alternative hypothesis H_1 (i.e., a difference between groups) and the null hypothesis H_0 (i.e., no difference between groups) [48]. For example, a Bayes factor of 3 would be interpreted as moderate evidence in favor of the alternative hypothesis and 40 would indicate very strong evidence.

submitted fully answered as no question was mandatory. Thus, we aimed to increase participants' willingness to state at least some extra information. Our report below includes all the available data. Figure 4 shows that participants continually responded to the ESQs.

The study participants started 126 learning sessions in the morning hours (4am – 12pm), 87 in the afternoon (12pm – 6pm), and 114 in the evening (6pm – 1am). There were no learning sessions between 1am and 4am. An overview of the learning sessions per hour is shown in Figure 3. The Google Activity Recognition API recognized 308 learning sessions as performed while *STILL*, five in a vehicle, two *STILL* and in a vehicle, two while walking, three *STILL* and walking, and finally, one case where all three activities occurred in the same learning session with a confidence above 90%, and six remained unknown. The majority of the sessions with ESQ data (236) took part at home and some outdoors (10), while traveling or commuting (9), at work (7). Four were specified as "other".

4.4.2 Characteristics of Interruptions (RQ2). We differentiate two types of interruptions: (1) interruptions that terminated the learning session and (2) interruptions that only led to a temporary suspension of the learning app. In the latter case, participants returned to the learning app within 10 minutes, the cut-off time after which we classified a learning session as ended (see Section 3.2). We first report the characteristics of the suspending interruptions and then continue with the terminating interruptions, or *termination events*. Below, unless otherwise indicated, "interruption" refers to "suspending interruption", while terminating interruptions are explicitly called "termination events".

Approximately 39% of learning sessions were interrupted and then continued within 10 minutes. During the 327 learning sessions, we recorded a total of 276 suspending interruptions. There were between 0 and 9 interruptions per session (M = 0.84, SD = 1.53) and an average interruption lasted 27.6 seconds (SD = 67.4s). The shortest suspending interruptions were barely a second long, the longest interruption was 8:46 minutes long. Of the 276 suspending interruptions, we classified 187 as internal (i.e., app switches without indication or without screen locks), 86 as device-internal (i.e., app switched due to calls, SMS, or notifications), and three as ambiguous (i.e., screen lock; see Table 1). This does not yet include the ESQ data, hence, there are no external interruptions reported. The interruptions our algorithm classified as device interruptions mostly followed a notification issued by a messaging app (52 of 86 cases, i.e., 60.5%).

Participants intentionally ended 168 of the 266 learning sessions supplemented with ES data, while 98 were not resumed after an interruption (36.8%). In 19 cases, participants confirmed that it was really necessary to interrupt the learning session (*Very important - it was urgent / time-critical*). For 38 situations, they selected a moderate level of urgency (*Moderate - I had to do it eventually in the near future*) and in 41 situations, the interruption was avoidable (*Not important - I could have ignored it and continued learning*). According to the ESQs, 37 learning sessions were terminated because of external, 33 because of internal, and 28 because of device-internal interruptions. In 62 cases, there was ESQ data available and our algorithm had detected a termination event, their types coincided in 20 cases, i.e., 32.2% (cf. Table 2). However, it has to be noted that it was not possible for us to uniquely identify internal termination events. In case the LAIRA app could not ultimately determine if the interruption was caused by an internal or external stimulus, the event was labeled "ambiguous" and later confirmed as "internal" or "external" through the ESQ. As Table 2 shows, the LAIRA app classified a total of 32 interruptions as device-internal that were later associated to external or internal stimuli by the participants.

Table 1: Clustered by the interruption type, this table presents an overview of the number of suspending interruptions in total as well as the minimum and maximum of interruptions in one learning session (more than one interruption possible). Further, the table outlines the length of these interruptions in seconds. Note that the type "ambiguous" contains interruptions where the automated classification could not ultimately determine if the source is internal or external.

Interruption Type	$\text{Count}_{\text{Total}}$	$\text{Count}_{\text{Max}}$	$\text{Count}_{\text{M (SD)}}$	$\text{Length}_{\text{M (SI}}$
Device-internal	86	6	0.26 (0.77)	51.94s (98.94
Internal	187	8	0.57 (1.20)	16.15s (38.83
Ambiguous	3	1	0.01 (0.01)	44.04s (66.77
Overall	276	9	0.84	27.65s (66.00

Table 2: This table concerns all terminating interruption events. We contrast the automated classification into device, internal, and ambiguous interruptions performed by the LAIRA app ("classified:", columns) with the subjective statements of participants as device, external, or internal through the ESQs ("ESQ:", rows).

	classified: ambiguous	classified: device	classified: internal
ESQ: device	0	15	6
ESQ: external	0	18	3
ESQ: internal	1	14	5
Total	1	47	14

Moreover, the comparison of ESQ data and the classification showed that of the 98 unintentionally ended learning sessions, our algorithm had associated 61 with a session termination event (61.2%). On the other hand, of the intentionally ended sessions, 73 were classified as device-internal (43.5%) and 51 as internal interruptions (30.3%). In 42 cases (25%), we detected no terminating interruption. Adding automatic classifications of termination events as shown in Figure 2 to the ESQ data, we arrive at a total of 37 external, 97 internal, 112 device interruptions, and 3 ambiguous (i.e., internal or external) interruptions.

4.4.3 Effect of interruption type on interruptions and termination risk (RQ3). The diverse nature of interrupting events and secondary tasks also manifests in characteristics of the interruptions, such as their duration. For example, we compared the duration of the detected suspending interruption types "internal" and "device-internal". The Bayes factor BF_{10} = 327.0 of a Bayesian ANOVA provides very strong evidence in favor of the hypothesis that their duration differed. Namely, the device-internal interruptions (N=86, M = 519.4s, SD = 989.4s) were typically longer than internal interruptions (N=197, M = 172.4s, SD = 440.0s). Moreover, when there was a suspending interruption in a learning session, participants were

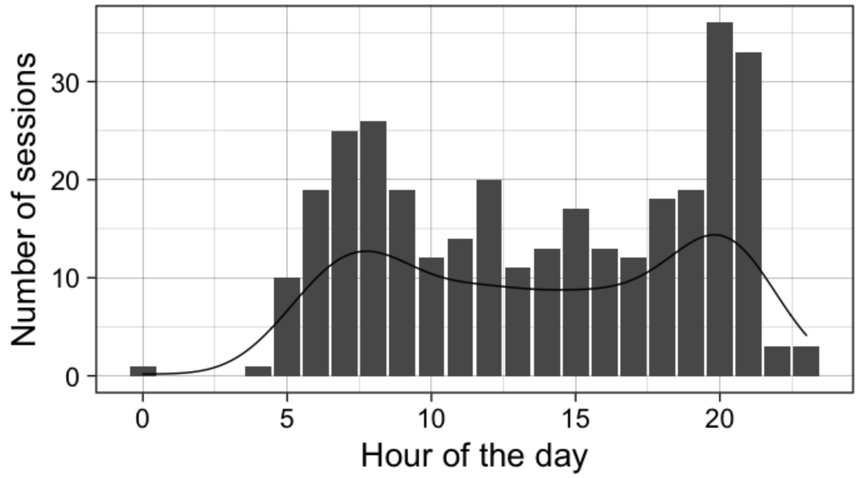


Figure 3: Learning session distribution over the day over all participants, and a scaled fitted distribution curve. Note the two peaks in the morning and evening, as well as a small increase in the number of sessions around lunchtime.

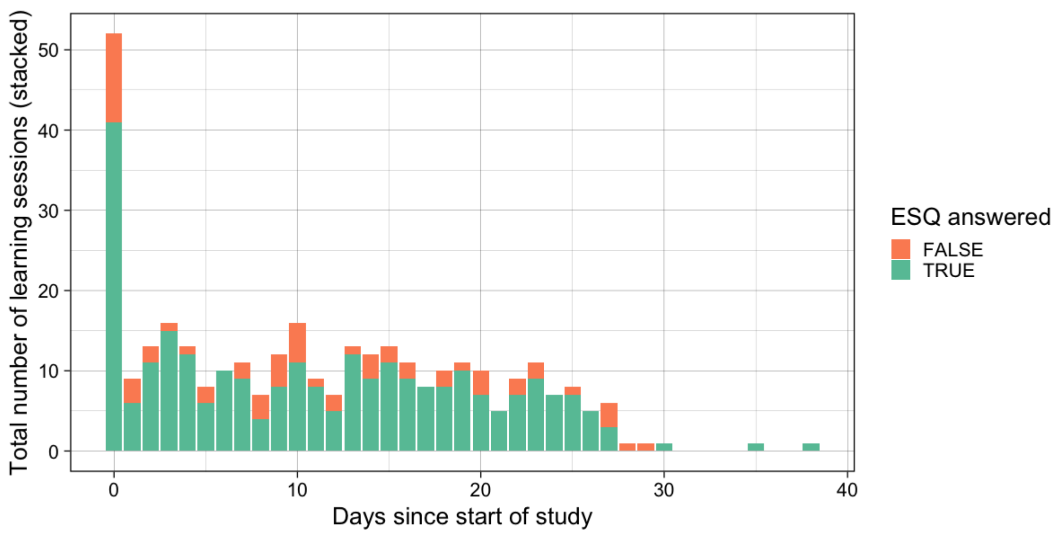


Figure 4: Total number of learning sessions per day of the study, aligned such that each participant's first learning session is on Day 0. Note that two participants logged additional sessions after the official end of the study.

more likely to end their learning session unintentionally than when the learning session was not interrupted (i.e., any response not equal to "I was done using the app" in the ESQ) ($\chi^2(1) = 10.913$, $p < 0.05$).

4.4.4 Effect of Interruptions on Learning Sessions (RQ4). We further found that the occurrence of suspending interruptions (yes|no) influenced the net length of learning sessions (in seconds *excluding* interruption time). The Bayes factor $BF_{10} = 89226.0$ shows very strong evidence for the duration of learning sessions with interruptions (N=128, $M = 877.6$s, $SD = 628.6$s) differing from learning sessions without interruptions (N=199, $M = 538.8$s, $SD = 422.6$s). Namely, learning sessions *with* interruptions were longer. An additional Spearman correlation analysis suggests a positive relationship between the total number of interruptions within sessions and the length of the learning session ($rs(327) = .284$, $p < .05$). In particular, we can see an increase in learning time with an increase in interruption counts (see Figure 5). Similarly, there was a significant positive correlation between the total length of all interruptions in a session and the length of the learning session ($rs(327) = .239$, $p < 0.05$). The rank-based tests were chosen because the assumption of normality was violated.

4.4.5 The Influence of Context on Learning Sessions (RQ5 - Part I). In order to analyze the role of context, we first modeled the

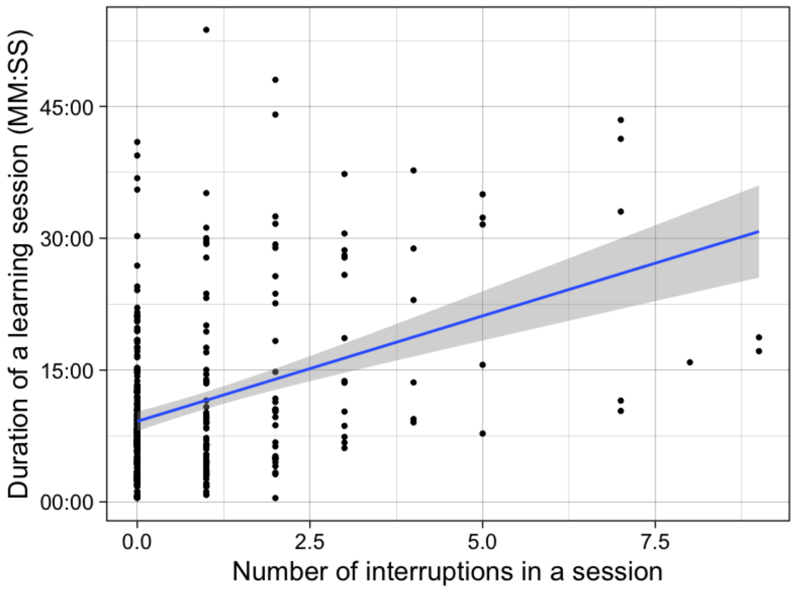


Figure 5: Correlation between the number of interruptions that occur in a learning session and the overall task time (excluding interruptions).

relationship between the time of day (morning, afternoon, evening), location (home, outdoors, work, travel or commute), and company (yes, no) and the resulting duration of learning sessions. We did not include movement data, as the number of samples for values other than "still" was low. Specifically, we constructed a GAMLSS with a Weibull distribution that predicted the duration of a learning session (*excluding* interruption time) based on the fixed factors above and the participants as random effect for all learning sessions with ESQ data. Here, we only present key points of the model. We provide the complete analysis in the supplementary material.

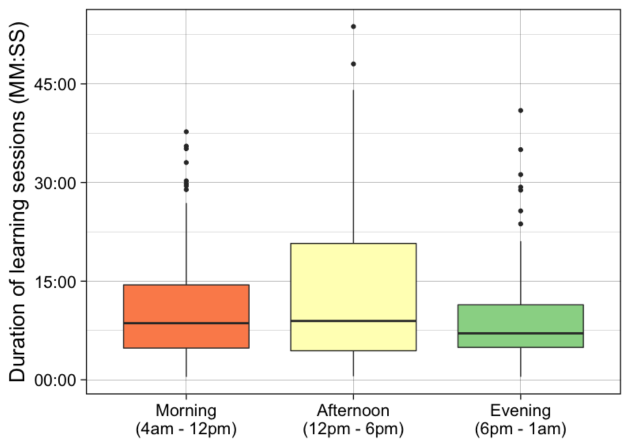
We found that, compared to learning sessions at home, sessions outdoors ($\beta = -0.41$, $SE = 0.21$, $t = -2.0$, $p < 0.05$) and while traveling or commuting ($\beta = -0.42$, $SE = 0.21$, $t = -2.0$, $p < 0.05$) led to shorter duration estimates. Afternoon sessions ($\beta = 0.07$) were estimated to be longer and evening sessions ($\beta = -0.05$) shorter than morning sessions but this effect was not significant (see also Figure 6a). Finally, having company led to a lower estimate for session length ($\beta = -0.23$, $SE = 0.09$, $t = -2.5$, $p < 0.05$). The intercept was significant at $p < 0.001$.

4.4.6 The Influence of Context on Interruptions (RQ5 - Part II). We also fitted GAMLSSs for the number of suspending interruptions and the total duration of all interruptions in one learning session. For interruption counts, we used a Poisson distribution to allow for 0 values.

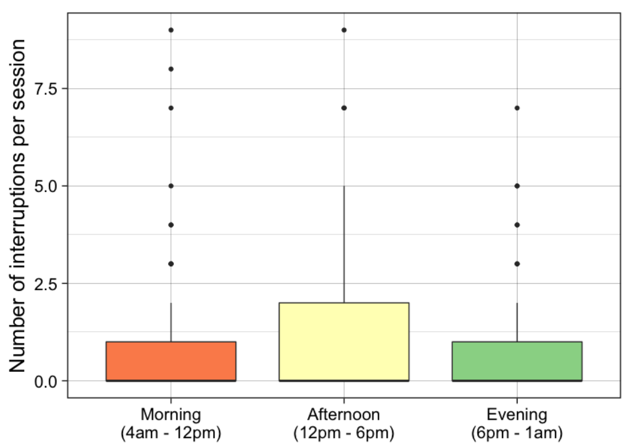
For the number of suspending interruptions, we set the environment, company, time of day, and triggered push notifications as fixed effects. Receiving notifications increased the predicted number of suspending interruptions ($\beta = 0.53$, $SE = 0.24$, $t = 2.22$, $p < 0.0.5$). Besides the intercept ($p < 0.05$), no other effects were significant. In particular, the occurrence of push notifications led to learners' interrupting their session to open a notifying app with a probability of 31.9% (thus causing a classification as device interruption). Additionally, the number of suspending interruptions at different times of the day is shown in Figure 6b.

Finally, we fit a model to predict the total duration of interruptions within a learning session from the same factors. The duration was estimated to be highest in the afternoon ($\beta = 0.69$, $SE = 0.29$, $t = 2.35$, $p < 0.05$ and while traveling or commuting, interruption time was shorter than at home ($\beta = -1.54$, $SE = 0.66$, $t = -2.35$, $p < 0.05$). The intercept was significant at $p < 0.001$.

4.4.7 Questionnaire: Subjective Assessments of Distractions & Termination. Our user study, in particular the ESQs, helped our participants to self-reflect on their learning behavior. To assess their subjective impression on how they handled interruptions during learning sessions, we presented them with a set of questions at the end of the study. Here, participants stated that they got most easily distracted by external interruptions ($M = 4.27$, $SD = 1.21$, 5-point Likert scale from 1="strongly disagree" to 5="strongly agree") compared to device ($M = 3$, $SD = 1.28$) and internal interruptions ($M = 3$, $SD = 1.48$). Four participants furthermore had the impression that they usually discontinued learning after an interruption occurred, while the majority (7) reported to usually continue learning after a short period of time. When asked how difficult it usually was for them to pick up a learning session after an interruption (5-point Likert scale from 1="very difficult" to 5="very easy"), participants stated medium difficulty, slightly leaning toward easy learning session resumption ($M = 3.45$, $SD = 1.3$). The people who reported problems with resuming the learning task further noted that the main reason was due to loss of focus (P3, P4) and not remembering "[...] what I was doing before the interruption" (P2). Those who did not find it difficult to continue a session after an interruption noted that they had established fixed habits (P1)



(a) **Duration of learning sessions**



(b) **Number of suspending interruptions**

Figure 6: Learning sessions and interruptions at different times of the day over all participants.

and that content in learning apps is fairly easy (P6, P10). One participant added that "with Duolingo, the sentences and questions do not necessarily need the context of previous exercises, so it was easy to continue a lesson without having a big disadvantage from the interruption" (P9).

5 DISCUSSION

5.1 Limitations

As our user study was conducted during the COVID-19 pandemic, the generalizability of our results is limited in regards to potentially different usage patterns of the application due to the anomalous daily routines of our participants during lockdown and work-from-home phases. Prior work suggests a wide variety of usage contexts for ML applications (e.g., [26, 42]), which we cannot confirm with the data we collected. The majority of the learning sessions we recorded show participants using learning apps at home (311) and only rarely on the go or in a vehicle (11). Still, participants experienced numerous interruptions, with one third of all learning sessions being disrupted. Considering that learning in a home setting is more controlled and quiet than what we can expect from outdoor or public spaces, we estimate that the number of interruptions could be even higher and more diverse for "normal" usage patterns. Similarly, with our sample being gender-biased (all female) and comparably small, a more diverse participant set would be needed to allow the generalization of learning and interruption patterns. Still, this study reveals interesting tendencies and potential patterns in mobile learning and provides implications for the design of ML in light of interruptions.

Furthermore, we annotated the logging data with supplementary information we gathered from the ESQs for the purpose of the study. In particular, we aimed to verify the cause of interruptions and identify the intentional termination of learning sessions. However, we do not know if the participants actually perceived (and remembered) the interruptions detected by LAIRA. Hence, it cannot be guaranteed that an ESQ response always matches the latest interruption. This would also partially explain the low coincidence rate of detected interruption types and the interruption causes selected in the ESQs. Moreover, the use of ESQs is not a feasible approach for everyday-use applications. Based on the data we collected, we can predict device-internal interruptions–caused predominantly by push notifications of messenger apps–but as of yet, fail to successfully distinguish between user-internal interruptions, which are not expressed by actions on the smartphone such as app switches, and external interruptions. The use of additional sensor data (e.g., to detect surrounding noise) or the application of machine learning to train a more sophisticated model of user behavior patterns could facilitate the better distinction of such interruptions.

5.2 Fragmented Use of ML Apps

Similar to prior work (e.g., [19, 37, 42]), we find that users' learning sessions vary in length. The median duration remains below 10 minutes (6:07 minutes), indicating a preference for short but frequent engagements with learning apps. Even in times of a pandemic, the participants in our study were not always at home and often in the company of other people when they studied with their ML app. This confirms previous findings on the ubiquity of mobile learning [13, 42]. We also found that the study environment as well as other contextual factors, namely the company of other people, received push notifications, and the time of day, at least partially explain the variation in the length of learning sessions and the length or number of interruptions within a learning session. In particular, almost 40% of the learning sessions were interrupted for up to nine minutes. Currently, many ML apps deal with short session times and frequent interruptions by splitting content into bite-sized chunks that can be studied independently. However, for many topics, such strategies are not suitable and it is a challenge to teach more complex concepts with ML apps that require longer periods of time and high levels of focus, e.g., in STEM subjects. In this work, we focus particularly on short interruptions during learning sessions. Hence, below, we discuss several strategies that mitigate the effects of interruptions and might be a first step towards teaching more complex concepts within ML apps.

5.3 Mitigation Potential

Our study revealed a great variety of interruptions that undoubtedly affect users during their mobile learning sessions. Due to this variety, not all interruptions can be detected automatically and mitigated using the same techniques. Subsequently, we describe four mitigation strategies, focusing on *avoiding* interruptions, *ignoring and postponing* them, preparing for upcoming interruptions, and supporting users in *resuming* the learning task after an interruption.

5.3.1 ***Avoiding*** *Interruptions.* The distribution of learning sessions across the day indicates increased usage in the morning and late evening hours (see Figure 3). We see two possible reasons for this pattern: (1) common work schedules that entail a dip after 8am and an increase at 5pm and (2) users' circadian rhythm. The circadian rhythm describes fluctuations of alertness and attention over the course of the day, indicating that cognitive and memory performance are highest around 2 hours and 12 hours after waking up, taking a dip in the time in between [12, 40]. Further, variations can occur because of people's individual times of productivity ("night owls" or "morning larks"). In our study, participants learned more often in the evening but session length increased in the afternoon compared to the evening. Furthermore, the likelihood of participants terminating their learning session after being interrupted was highest in the afternoon. Prior work by Schoedel et al. [43] presented the automatic classification of activity behavior based on phone logging data. Similar approaches could be used to recommend individually optimal moments for ML sessions. Hereby, the aim would be to maximize the learners' level of attention and focus by reducing the likelihood of interruptions and the frequency of users succumbing to them.

5.3.2 ***Ignoring*** *or* ***Postponing*** *Interruptions.* When LAIRA detected a device interruption, the most frequent source was an incoming push notification. When a notification occurred, participants reacted to this notification in one third of the cases and, thus, interrupted their learning session. In most cases, participants switched to messaging applications. This means that notification management systems (e.g., [16, 20, 31, 34]) that defer notifications until an activity breakpoint is detected would also be a promising approach for mitigating interruptions in ML. This approach is further supported by the participants' impression that nearly 80% of the interruptions could be either postponed or even ignored. Context variables such as those we gathered with the activity logging and ESQs could hint at suitable moments for presenting notifications [2].

5.3.3 ***Preparing*** *for Upcoming Interruptions.* If an interruption is detectable but not avoidable, at the very least, the learner could be guided to the end of a learning unit and be prepared for an upcoming interruption [9, 47]. Prior work investigated the exploitation of the interruption lag, the time window between noticing an interruption and actually switching to the secondary task. This short time window provides the opportunity to mentally prepare, for example, by consolidating the memory of what a user was currently doing or displaying what they were about to do next [1]. This can be achieved by presenting a summary of what the user just learned or by suggesting them to take written or mental notes to help resume the learning task later on (cf. [17]).

5.3.4 ***Resuming*** *Learning after Interruptions.* The data gathered in our study show that the occurrence of interruptions affects learning sessions. We found that sessions that were suspended by an interruption are longer (even after subtracting the interruption time). Furthermore, the net overall session time increases with the number of interruptions. This goes in line with prior work suggesting that it takes time to resume a task after an interruption (called the "resumption lag" [47]) and that overall task completion time can increase [3, 22, 25]. These results indicate the need for technology interventions to minimize the resumption lag by guiding users back to the original task after the interruption has passed. Task resumption support has shown to positively influence task completion time and error rate after an interruption [39].

6 CONCLUSION

This paper presents an investigation of the usage of mobile learning apps as well as the origins and effects of interruptions. By deploying the LAIRA application in a four-week field study, we were able to gather information on the users' mobile learning behavior and various occurring interruptions in the wild. Our analysis shows that interruptions during ML are frequent, with 276 recorded interruptions in 327 learning sessions, and are caused by a variety of internal and external stimuli. Although many of the occurring interruptions were comparably short, more than one third of learning sessions were terminated because of an interruption. More importantly, participants reported that only 20% of the interruptions were urgent, while all other interruptions could have been ignored or postponed. This opens up a great number of opportunities to design strategies to mitigate interruptions during mobile learning –including avoiding and managing interruptions, preparing for them, and supporting the resumption of the learning task after an interruption. Future work is required to evaluate the proposed strategies with respect to their effectiveness in mitigating ML interruptions. By tackling the disruptive effects of interruptions, mobile learning applications have the potential to extend their teaching capabilities by allowing for more focused learning.

ACKNOWLEDGMENTS

This research is financially supported by the European Union's Horizon 2020 Programme under ERCEA grant no. 683008 AMPLIFY and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

REFERENCES

[1] Erik M Altmann and J Gregory Trafton. 2002. Memory for goals: An activation-based model. *Cognitive science* 26, 1 (2002), 39–83.

[2] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A Survey of Attention Management Systems in Ubiquitous Computing Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (July 2018), 1–27. https://doi.org/10.1145/3214261

[3] Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708. https://doi.org/10.1016/j.chb.2005.12.009

[4] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (Dec. 2017), 1–40. https://doi.org/10.1145/3123988

[5] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR)* 50, 6 (2017), 1–40.

[6] Cindy Blanco. 2020. 2020 Duolingo Language Report: Global Overview. https://blog.duolingo.com/global-language-report-2020/

[7] Shadan Sadeghian Borojeni, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2016. Peripheral Light Cues for In-Vehicle Task Resumption. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM, Gothenburg Sweden, 1–4. https://doi.org/10.1145/2971485.2971498

[8] Peter A Bruck, Luvai Motiwalla, and Florian Foerster. 2012. Mobile Learning with Micro-content: A Framework and Evaluation., In Bled eConference. *Bled eConference* 25, 527–543.

[9] Duncan P Brumby, Anna L Cox, Jonathan Back, and Sandy JJ Gould. 2013. Recovering from an interruption: Investigating speed-accuracy trade-offs in task resumption behavior. *Journal of Experimental Psychology: Applied* 19, 2 (2013), 95.

[10] Fabio Ciravegna. 2019. Creating a never ending background service in Android > 7. https://fabcirablog.weebly.com/blog/creating-a-never-ending-background-service-in-android-gt-7

[11] William L Cull. 2000. Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology* 14, 3 (2000), 215–235.

[12] C. A. Czeisler and J. J. Gooley. 2007. Sleep and Circadian Rhythms in Humans. *Cold Spring Harbor Symposia on Quantitative Biology* 72, 1 (Jan. 2007), 579–597. https://doi.org/10.1101/sqb.2007.72.064

[13] Valérie Demouy, Ann Jones, Qian Kan, Agnes Kukulska-Hulme, and Annie Eardley. 2016. Why and How Do Distance Learners Use Mobile Devices for Language Learning? *The EuroCALL Review* 24, 1 (2016), 10–24.

[14] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language learning on-the-go: opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17*. ACM Press, Vienna, Austria, 1–12. https://doi.org/10.1145/3098279.3098565

[15] Fiona Draxler, Christina Schneegass, and Evangelos Niforatos. 2019. Designing for Task Resumption Support in Mobile Learning. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Taipei Taiwan, 1–6. https://doi.org/10.1145/3338286.3344394

[16] Joel E. Fischer, Chris Greenhalgh, and Steve Benford. 2011. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. ACM Press, Stockholm, Sweden, 181. https://doi.org/10.1145/2037373.2037402

[17] Victor M. González and Gloria Mark. 2004. "Constant, constant, multi-tasking craziness": managing multiple working spheres. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*. ACM Press, Vienna, Austria, 113–120. https://doi.org/10.1145/985692.985707

[18] Helen M Hodgetts and Dylan M Jones. 2006. Interruption of the Tower of London task: support for a goal-activation approach. *Journal of Experimental Psychology: General* 135, 1 (2006), 103.

[19] Theo Hug and Martin Lindner (Eds.). 2006. *Micromedia & e-Learning 2.0: Gaining the Big Picture. Proceedings of Microlearning Conference 2006* (1. ed ed.). Innsbruck Univ. Press, Innsbruck. OCLC: 699639536.

[20] Shamsi T. Iqbal and Brian P. Bailey. 2007. Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*. ACM Press, San Jose, California, USA, 697–706. https://doi.org/10.1145/1240624.1240732

[21] Shamsi T. Iqbal and Eric Horvitz. 2007. Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, San Jose, California, USA, 677. https://doi.org/10.1145/1240624.1240730

[22] Ioanna Katidioti, Jelmer P Borst, Marieke K van Vugt, and Niels A Taatgen. 2016. Interrupt me: External interruptions are less disruptive than self-interruptions. *Computers in Human Behavior* 63 (2016), 906–915.

[23] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 4521–4532. https://doi.org/10.1145/2858036.2858465

[24] Dagmar Kern, Paul Marshall, and Albrecht Schmidt. 2010. Gazemarks: gaze-based visual placeholders to ease attention switching. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 2093. https://doi.org/10.1145/1753326.1753646

[25] John G Kreifeldt and ME McCarthy. 1981. Interruption as a test of the user-computer interface. In *Proceedings of the 17th Annual Conference on Manual Interaction*. 655–667.

[26] Agnes Kukulska-Hulme. 2009. Will mobile learning change language learning? *ReCALL* 21, 02 (May 2009), 157. https://doi.org/10.1017/S0958344009000202

[27] Alexander Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N. Patel, and Jacob O. Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 2953–2962. https://doi.org/10.1145/2702123.2702539

[28] Mark A McDaniel, Gilles O Einstein, Thomas Graham, and Erica Rall. 2004. Delaying execution of intentions: Overcoming the costs of interruptions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 18, 5 (2004), 533–547.

[29] Jessica S. Mendoza, Benjamin C. Pody, Seungyeon Lee, Minsung Kim, and Ian M. McDonough. 2018. The effect of cellphones on attention and learning: The influences of time, distraction, and nomophobia. *Computers in Human Behavior* 86 (2018), 52 – 60. https://doi.org/10.1016/j.chb.2018.04.027

[30] Christopher A Monk, J Gregory Trafton, and Deborah A Boehm-Davis. 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of experimental psychology: Applied* 14, 4 (2008), 299.

[31] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. ACM Press, Osaka, Japan, 475–486. https://doi.org/10.1145/2750858.2807517

[32] Antti Oulasvirta. 2005. Interrupted cognition and design for non-disruptiveness: the skilled memory approach. In *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*. ACM Press, Portland, OR, USA, 1124. https://doi.org/10.1145/1056808.1056843

[33] Antti Oulasvirta and Pertti Saariluoma. 2006. Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies* 64, 10 (Oct. 2006), 941–961. https://doi.org/10.1016/j.ijhcs.2006.04.006

[34] Romina Poguntke, Christina Schneegass, Lucas Van der Vekens, Rufat Rzayev, Jonas Auda, Stefan Schneegass, and Albrecht Schmidt. 2020. NotiModes: an investigation of notification delay modes and their effects on smartphone users. In *Proceedings of the Conference on Mensch und Computer*. ACM, Magdeburg Germany, 415–419. https://doi.org/10.1145/3404983.3410006

[35] R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

[36] R. A. Rigby and D. M. Stasinopoulos. 2005. Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics* 54 (2005), 507–554.

[37] Fernando Rosell-Aguilar. 2018. Autonomous language learning through a mobile application: a user evaluation of the *busuu* app. *Computer Assisted Language Learning* 31, 8 (Nov. 2018), 854–881. https://doi.org/10.1080/09588221.2018.1456465

[38] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. 2014. Large-scale Assessment of Mobile Notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 3055–3064. https://doi.org/10.1145/2556288.2557189

[39] Farzan Sasangohar, Stacey D Scott, and Mary L Cummings. 2014. Supervisory-level interruption recovery in time-critical control tasks. *Applied ergonomics* 45, 4 (2014), 1148–1156.

[40] Christina Schmidt, Fabienne Collette, Christian Cajochen, and Philippe Peigneux. 2007. A time to think: circadian rhythms in human cognition. *Cognitive neuropsychology* 24, 7 (2007), 755–789.

[41] Christina Schneegass and Fiona Draxler. 2021. Designing Task Resumption Cues for Interruptions in Mobile Learning Scenarios. In *Technology-Augmented Perception and Cognition*, Tilman Dingler and Evangelos Niforatos (Eds.). Springer International Publishing, Cham, 125–181. https://doi.org/10.1007/978-3-030-30457-7_5 Series Title: Human–Computer Interaction Series.

[42] Christina Schneegass, Nađa Terzimehić, Mariam Nettah, and Stefan Schneegass. 2018. Informing the Design of User-adaptive Mobile Language Learning Applications. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia - MUM 2018*. ACM Press, Cairo, Egypt, 233–238. https://doi.org/10.1145/3282894.3282926

[43] Ramona Schoedel, Florian Pargent, Quay Au, Sarah Theres Völkel, Tobias Schuwerk, Markus Bühner, and Clemens Stachl. 2020. To Challenge the Morning Lark and the Night Owl: Using Smartphone Sensing Data to Investigate Day–Night Behaviour Patterns. *European Journal of Personality* 34, 5 (Sept. 2020), 733–752. https://doi.org/10.1002/per.2258

[44] Stacey D. Scott, Stéphane Mercier, M. L. Cummings, and Enlie Wang. 2006. Assisting interruption recovery in supervisory control of multiple UAVs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. SAGE Publications Sage CA: Los Angeles, CA, 699–703. https://doi.org/10.1177/154193120605000518

[45] Mikis D. Stasinopoulos (Ed.). 2017. *Flexible regression and smoothing: using GAMLSS in R.* CRC Press/Taylor & Francis Group, Boca Raton.

[46] Peyman Toreini, Moritz Langner, and Alexander Maedche. 2018. Use of attentive information dashboards to support task resumption in working environments. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ACM, Warsaw Poland, 1–3. https://doi.org/10.1145/3204493.3208348

[47] J Gregory Trafton, Erik M Altmann, Derek P Brock, and Farilee E Mintz. 2003. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies* 58, 5 (2003), 583–603.

[48] Eric-Jan Wagenmakers, Jonathon Love, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Ravi Selker, Quentin F. Gronau, Damian Dropmann, Bruno Boutin, Frans Meerhoff, Patrick Knight, Akash Raj, Erik-Jan van Kesteren, Johnny van Doorn, Martin Šmíra, Sacha Epskamp, Alexander Etz, Dora Matzke, Tim de Jong, Don van den Bergh, Alexandra Sarafoglou, Helen Steingroever, Koen Derks, Jeffrey N. Rouder, and Richard D. Morey. 2018. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review* 25, 1 (Feb. 2018), 58–76. https://doi.org/10.3758/s13423-017-1323-7

[49] W.L. Yeung and Simon Y.W. Li. 2016. Prototyping the Machine-Human Dialogues in a Smartphone Voice Call Application With Task Resumption Support. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, Santa Clara, California, USA, 1788–1793. https://doi.org/10.1145/2851581.2892464