JITAI-Wizard: An Open-Source Android and Wear OS Tool for Wearable Just-in-Time Adaptive Interventions in Wizard-of-Oz Studies

Robin Burchard, Sebastian Jens Koschorreck, Jason Alexander Klaas, and Kristof Van Laerhoven

University of Siegen, 57076 Siegen, Germany
robin.burchard@uni-siegen.de

Abstract. Just-in-time adaptive interventions (JITAIs) deliver support exactly when people need it most, using sensor and context information to time short interventions around moments of heightened risk or opportunity. They are especially promising for wearable devices, which can provide always-available, low-friction prompts in everyday life. Yet early-stage JITAI research is slowed by the engineering effort required to build full sensing and decision pipelines before basic questions about timing, content, means of gaining user attention, and user experience factors can be explored.

We present JITAI-Wizard, an open-source[1] Android and WearOS framework for conducting Wizard-of-Oz studies of wearable JITAIs. In our approach, a human operator remotely triggers predefined micro-interventions manually or automatically on a smartwatch, simulating an intelligent JITAI. Our free and fully open-source software allows the community to design and evaluate interventions, explore use cases, and test hypotheses. The Wizard-of-Oz element is further strengthened by a programmable, minimalistic avatar, acting as a smart assistant. To validate our software, we conducted JITAI experiments with 10 participants in an induced intensive hand-washing setting, where participants with previously dirtied hands received interventions when washing their hands for too long. We evaluated the response rates, hand-washing durations, and general user feedback, and concluded how JITAI design can profit from our openly available application framework.

Keywords: Wear OS · JITAI · Wearable Sensing · Just-in-Time Adaptive Interventions · Wizard-of-Oz

1 Introduction

Wearable devices have become a pervasive platform for sensing and interaction, enabling continuous capture of physiological, behavioral, and contextual signals in everyday environments. In parallel, advances in human activity recognition (HAR) have made it possible to infer fine-grained user states such as physical

[1] `https://github.com/RBurchard/JITAI-WizardAndroid`

activity, routines, and context transitions from multi-modal sensor streams [29]. This convergence has created new opportunities for delivering timely, personalized interventions directly on the body, without requiring explicit user input or attention shifts to secondary devices [1].

Just-in-time adaptive interventions (JITAIs) operationalize this vision by delivering support at moments of heightened relevance, such as when a user is at risk of performing an undesired behavior or is particularly receptive to guidance [17]. They have been shown to have moderate to large effect sizes in health-related studies [26]. In contrast to static or schedule-based notifications, JITAIs rely on continuous sensing, inference, and decision-making pipelines to determine *when* and *how* to intervene. For wearable platforms such as smartwatches, this paradigm is particularly attractive: interventions can be short, unobtrusive, and tightly integrated into ongoing activities. As an example, Wang et al. identified just-in-time support as a promising approach for helping individuals with obsessive-compulsive disorder (OCD) manage symptoms in everyday life [27], highlighting opportunities for personalized symptom tracking and context-aware interventions that address challenges not adequately supported by existing treatment tools.

Despite this promise, the development of JITAIs remains technically demanding. Implementing a fully automated pipeline requires robust activity recognition models, reliable context detection, and carefully tuned decision logic. In early research stages, these components are often underspecified or subject to rapid iteration. As a result, substantial engineering effort is invested before fundamental questions can be addressed, such as which intervention timing strategies are effective, how one can best gain the user's attention, how users perceive interruptions in situ, or how intervention content should be structured. This creates a bottleneck that slows down exploratory studies and limits the ability to iterate on intervention design.

Wizard-of-Oz (WoZ) methodologies offer a pragmatic alternative by decoupling intervention design from automation. In a WoZ setup, a human operator simulates parts of the intelligent system, allowing researchers to study user interaction and system behavior without requiring a fully implemented backend. While WoZ approaches are well established in human-computer interaction [22], their application to wearable JITAIs remains underexplored, and existing solutions are often ad hoc, non-reproducible, or tightly coupled to specific study designs.

In this work, we address this gap by introducing *JITAI-Wizard*, an open-source framework for conducting Wizard-of-Oz studies of wearable JITAIs on Android and WearOS. The system enables a remote operator to trigger predefined interventions on a smartwatch in real time, simulating an intelligent intervention pipeline. By abstracting away sensing and inference components, JITAI-Wizard allows researchers to focus on intervention timing, content, and user experience, while maintaining ecological validity in real-world settings. The framework is designed to be modular, extensible, and easy to deploy. By allowing researchers to define experiment structure flows and intervention specifications with an intuitive graphical smartphone user interface and a completely optional multiplatform desktop application, we lower the barrier to entry for researchers in both HAR and the digital health domain. We explicitly target researchers with no background in programming or computer science, for whom the development and evaluation of such JITAI experiments is made easy by JITAI-Wizard.

To demonstrate the feasibility and utility of our approach, we evaluate JITAI-Wizard in an application scenario. We study intervention timing in a hand-washing context, where participants receive prompts when washing their hands for extended durations. The feasibility demonstration alsa o serves as real-world validation test for the stability and usability of the developed system.

The contributions of this work can be summarized as follows:

1. Creation and making available of an open-source Android and WearOS app framework for creating and delivering wearable just-in-time adaptive interventions
2. Design and implementation of various interventions, with and without attached micro-actions and mini-games
3. Evaluation of the created toolkit and intervention in an exemplary user study with regard to induced enacted compulsive hand-washing

2 Related Work

Our work builds upon foundational research in Just-in-Time Adaptive Interventions (JITAIs), mobile and wearable sensing frameworks, and the application of Wizard-of-Oz (WoZ) methodologies within ubiquitous computing.

2.1 Just-in-Time Adaptive Interventions (JITAIs)

Just-in-Time Adaptive Interventions (JITAIs) utilize continuous data streams to deliver behavior change support at optimal moments of vulnerability or receptivity [18]. Prior work has demonstrated the efficacy of JITAIs across diverse clinical and health contexts, including physical activity deployment [11], addiction recovery management [7], and weight loss [19]. A work by Mishra et al. [16] showcases machine-learning-based receptivity modeling for interventions. On wearable devices, micro-interventions leverage low-friction modalities to minimize user burden and cognitive interruption [8]. However, maximizing the efficacy of these prompts requires precise configuration of interaction dynamics, contextual triggers, and user receptivity models [21,12,16]. Because optimizing these discrete variables using fully automated pipelines requires significant, early-stage engineering, researchers face structural bottlenecks when iterating on initial design choices or evaluating foundational user experience dynamics [10,13].

2.2 Mobile and Wearable Sensing Toolkits

To reduce the engineering overhead of building context-aware software, the ubiquitous computing community has developed several extensible, open-source sensing frameworks. Platforms such as AWARE [5], Sensus [28], and Ohmage [23] offer robust abstractions for background sensor data aggregation, storage, and remote telemetry. While these frameworks often feature notification engines capable of triggering alerts based on raw sensor thresholds or scheduled rules, they are fundamentally optimized for passive data harvesting rather than interactive, human-in-the-loop intervention loops. Crucially, configuring these toolkits typically demands advanced software engineering or data pipeline proficiency, creating a technical barrier for non-technical domain experts, clinical researchers, and behavioral scientists [15]. *JITAI-Wizard* addresses this capability gap by decoupling context logic from data collection, offering a codeless operational layer specifically tailored for deploying interactive JITAI experiment scripts.

2.3 Wizard-of-Oz (WoZ) in Ubiquitous Computing

While the Wizard-of-Oz (WoZ) methodology is often used in robotics [20,6], it is also a core human-computer interaction technique used to evaluate user experiences with highly complex, simulated, or unbuilt intelligent systems [3,9]. Within mobile and ubiquitous computing, WoZ infrastructure has been deployed to prototype location-aware applications [4] and to the capturing of naturalistic mobile notification receptivity data [2]. Despite these applications, historical WoZ deployments in mobile environments are overwhelmingly characterized by ad-hoc, proprietary architectures built for singular study designs [14]. These bespoke solutions rarely support native and flexible integration with commercial smartwatch ecosystems like Wear OS, limiting ecological validity during real-world, in-the-wild evaluations. By providing a structured, open-source framework targeting wearable configurations, *JITAI-Wizard* establishes a reproducible mechanism for prototyping micro-interventions without manual backend development.

3 Methods

For this work, we develop Android and Wear OS applications, along with an optional desktop application, for sending configurable, adaptive just-in-time interventions and evaluate them in a user study with 10 participants. In this section, we first discuss the implementation details and design choices. Afterwards, we describe the evaluation with the user study and the experiments we conducted.

3.1 Interventions

As discussed in Section 2, different intervention styles, means of intervention delivery, and gaining the user's attention exist.

We focus on two modalities for gaining the user's attention: Vibration of the smartwatch and an alert sound played on the smartwatch. These are simple to deliver, as they can be triggered using built-in device functionalities and hardware (speaker and linear vibration motor). Additionally, they are commonly known as notification indicators to the user. In addition to either only playing

Table 1: Intervention types and mini-games that are currently available for selection. Interventions can be combined with an optional mini-game to keep the user's attention and distract them from their previous activity.

Intervention / Mini-game	**Type**	**Description**	**Parameters**
Text	Intervention	A simple text is displayed together with an "OK" button.	Text to display
Timer	Intervention	A text is displayed and a timer counts downwards, with an "OK" button.	Text to display, duration
Yes / No question	Intervention	A short text is displayed together with Yes/No buttons.	Question to display
Simon Says	Mini-game	Four colored squares light up one after the other and the user repeats the sequence.	Maximum length of sequence
Stand Still	Mini-game	The smartwatch must not be moved for x seconds (progress is displayed).	Duration in s
Lock Picking	Mini-game	Swiping in a circle, then tapping a button while a moving dot is in a highlighted area.	–
Trivia Question	Mini-game	A question game with up to four answers.	Questions with 3 wrong and 1 correct answers

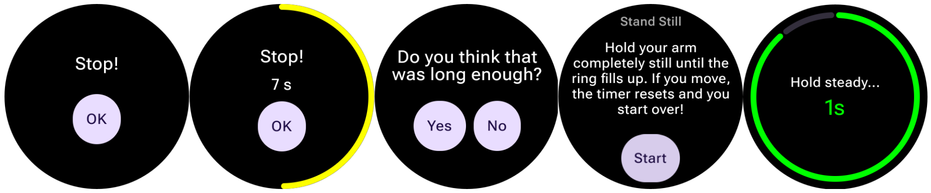


Fig. 1: Screenshots of the WearOS application. Left to right: Text intervention, Timer intervention, Yes/No question intervention, mini-game instructions for “Stand still”, and “Stand still” mini-game.

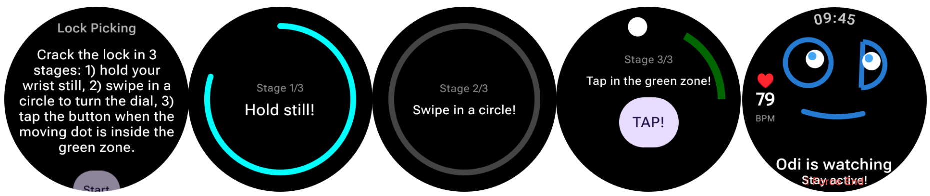


Fig. 2: Screenshots of the WearOS application. Left to right: Mini-game instructions for “Lock picking”, stages 1-3 of the “Lock picking” mini-game, main activity screen of the WearOS application, featuring the digital assistant avatar “Odi”.

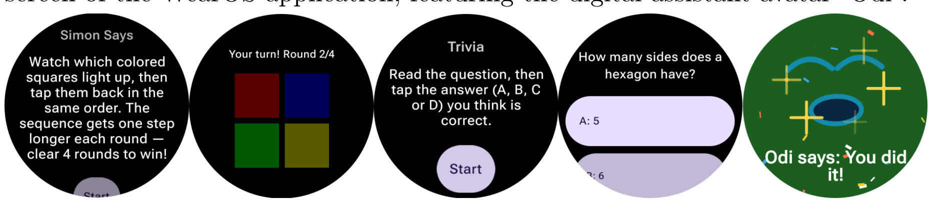


Fig. 3: Screenshots of the WearOS application. Left to right: Mini-game instructions for “Simon Says”, “Simon Says” mini-game, mini-game instructions for “Trivia”, “Trivia” mini-game, success/celebration screen shown to the user after finishing a mini-game, featuring “Odi”.

a sound or only letting the smartwatch vibrate, we also test a combination of vibration and alert sound.

We implement and offer different intervention types, as listed in Table 1. The table shows all intervention types and mini-games that can currently be selected. Example screenshots of these interventions and mini-games are shown in Fig. 1, Fig. 2, and Fig. 3. Informational interventions can inform the user of certain facts. E.g., the smartwatch displays a text like “you have been washing long enough”. A simple text intervention could be used as well, e.g., “Stop” could be displayed. Other options are timers (“when the timer ends, you should really stop”), feedback (showing the current duration of an activity compared to a desirable duration), or exercises (to divert the user to a different, non-harmful activity). The exercises are mini-games, such as a “lock picking” task, “holding still”, a “Simon says” game, and even fully customizable trivia questions.

All the intervention types are available and can be freely combined or selected in the experiment applications. Durations can be pre-selected or manually and dynamically controlled via the smartwatch and desktop applications.

The application allows the user to create fully customizable experiment scripts, which can be exported and imported to a simple JSON format if desired. The experiment can then be run with full logging capabilities, so that each step in the experiment can later be retraced from the logged information in the session file.

This broad range of abilities ensures that the use cases of a potential future JITAI study are covered by the available implementation. However, for our validation study, we could not test all intervention types and combinations of parameters, as this would have required a larger number of participants. However, we evaluated the most promising combinations in a user study, as explained in Section 3.3.

3.2 Implementation and Information Flow

The underlying idea of our software framework is to split the software into a smartwatch application that delivers interventions and a smartphone application that controls and remotely triggers them. Optionally, the smartphone, and through it, the smartwatch can be controlled wirelessly from a desktop application, which comes with an intuitive graphical user interface. By controlling the smartwatch from the phone, a human expert or experimenter can decide how and when to deliver interventions. If needed, these interventions can be triggered secretly, without letting the user know that the smartwatch is controlled remotely, turning the experiment into a Wizard-of-Oz study.

Software Structure We split our software into three packages: *jitaiconvention*, an Android package with all shared classes, *jitaiphone*, the Android phone app for controlling the experiment and triggering interventions, and *jitaiwear*, the WearOS application that we use to deliver the interventions. We wrote all Android and WearOS code in Kotlin using Android Studio. For the UI, we employ Jetpack Compose. The desktop application is written in C# using .NET and Avalonia UI[2], and supports Windows, macOS, and Linux.

System Overview In Fig. 4, we display the overview of our system. Note that the Desktop application is entirely optional. While it is probably easier to edit, save, and load the experiment scripts via the desktop application, the entire framework can be used without a desktop pc, with only a smartphone and a smartwatch.

Convention Library The convention package is used to store all shared information between the two applications. Currently, this includes an "intervention" class, which defines all properties of an intervention that should be triggered, as

[2] `https://avaloniaui.net`

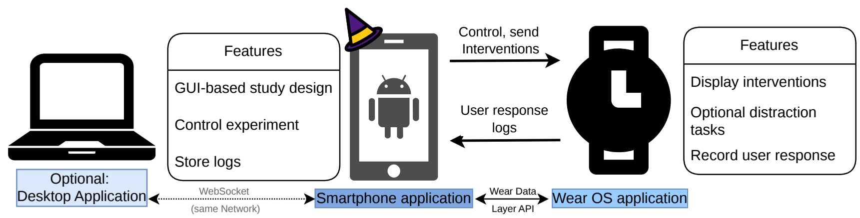


Fig. 4: Overview of devices, features, and information flow. The smartphone application controls the experiments and can be used to pre-design study procedures. While conducting an experiment, it additionally stores all feedback from the smartwatch in an experiment log file. The desktop application can optionally be used to remotely control the experiments. The WearOS application displays the interventions and distraction tasks, and records the user's responses and timings.

well as a "protocol" class, which defines the communication interface and paths to use between phone and smartwatch. For triggering the interventions and all other inter-device communication, the Wear Data Layer API[3] is used. Although this API supports sending messages via servers owned by Google, this is deactivated by default if no Google accounts are installed on the devices or if no internet connection is made available. Instead, the communication is handled locally, and the transmissions are sent and received via Bluetooth.

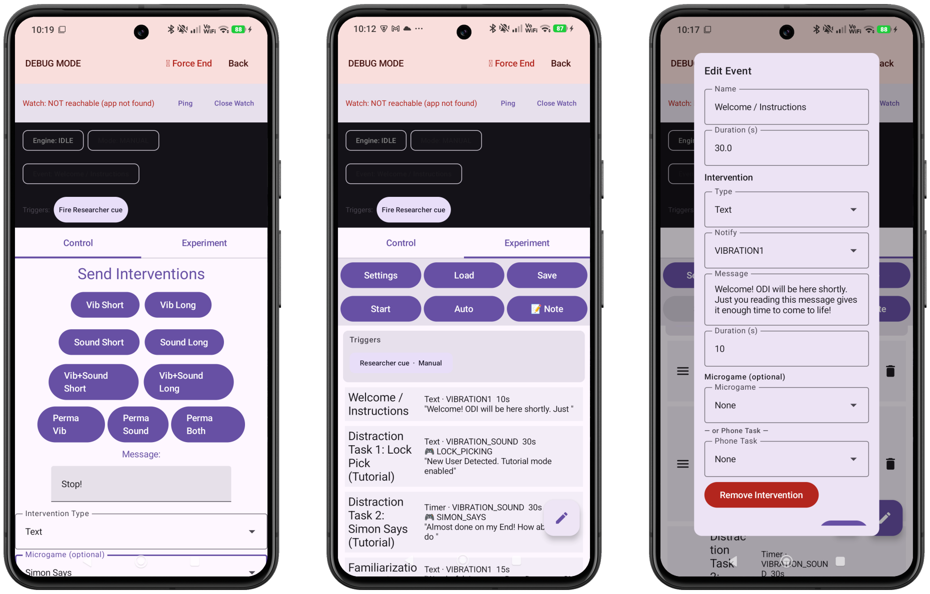
Android Phone App We built our phone app to support different use cases, from a simple manual sending of interventions to complex experiment structures with logging, timers, and full customization. Screenshots of the app are displayed in Fig. 5.

To account for the simplest version of "just" delivering interventions, we created a control panel view, which displays all necessary settings for the intervention and has a button to trigger an intervention according to the current settings. This view also contains a button to cancel any ongoing interventions.

Additionally, we added a complex experiment flow view to the phone application, from which it is possible to set up, restructure, save, and load an arbitrary experiment script. The experiment script always consists of a list of steps to be undertaken. This list can easily be created or altered by any non-expert user via the touchscreen of the phone. It is then possible to "start" an experiment script on the screen. Once the experiment is started, the phone will show the steps of the experiment and buttons for transitioning to the next step. Wherever the script contains the sending of interventions, buttons are displayed to administer these interventions.

The app also contains a general settings page that can be used to define global settings, like keeping logs and naming patterns for log files.

[3] `https://developer.android.com/training/wearables/data/overview`



(a) Control panel view. (b) Complex experiment. (c) Edit mode.

Fig. 5: Screenshots from the accompanying Android phone application (in debug mode): (a) Sending single interventions manually: Settings for a single intervention can be changed, including the text to display, the type of intervention, the attached mini-game, and the duration of the intervention. The intervention is sent by pressing a button with the desired notification sound/vibration pattern. (b) Complex experiment view: Via the buttons at the top, one can save, load, start, or stop entire experiment structures. The list of steps is displayed below. With a tap on the pencil button in the lower right corner, edit mode is activated. (c) Edit mode, editing an event. The edit mode is active, and one can edit an event with title, duration, and the optional intervention to be sent to the users. Note that "phone task" refers to a BETA feature not yet available in the app.

WearOS app In contrast to the feature-heavy smartphone app, the WearOS app for the smartwatch is lightweight and simple. Its main activity shows the user's current BPM, and our digital avatar, "Odi" (or "ODI", cf. Fig. 2), but the interventions will only be displayed when triggered via the smartphone application. Odi can also be used to deliver messages to the user, acting as an "AI" assistant. Currently, it is of course controlled manually by the researcher. The WearOS app itself registers a listener for the Wear Data Layer API, and can launch itself from the background, without declaring or running a background or foreground service, which minimizes its impact on the watch's battery life.

Once an intervention is triggered from the smartphone, the watch's screen will turn on, it will vibrate and/or play a sound according to the specification, and then it will show the selected intervention. Depending on the type of intervention, a different intervention screen is displayed, either with a certain text or with a timer or an exercise prompt to follow. If configured, the user is then tasked with solving a mini-game distraction task. The user can also confirm the intervention with a press of a button, or the experimenter can cancel the intervention using the smartphone app. Additionally, a lifespan can be set for the intervention, after which it auto-closes itself.

Desktop Application The desktop application is built in C# using .NET and Avalonia UI. It is optional, but can be used to plan and edit an experiment script as well as to run it. The experiment data is synced between the smartwatch and the desktop application if they are connected. Additionally, the smartwatch can be configured to continuously stream health and movement-related sensor data to the desktop application via the smartphone. The desktop application offers some additional features (live display of device rotation, recording of sensor data), but the smartwatch application remains standalone, so that experiments can be planned, conducted, and logged without the requirement of a connected desktop device.

The availability of the Desktop application makes it even easier to pre-plan experiment scripts and to export the recorded logs and feedback directly on a Desktop device.

Open source policy and license The complete source code is free and openly available under the MIT License at GitHub[4][5], along with instructions on how to build and install the applications. Overall, the application is designed to cover as many experimental use-cases in the JITAI experiment domain as possible, as well as to be easy to extend in code. See Section 5 for the complete Code and Data availability statement.

3.3 Experiments

We conducted a user study with 10 participants to evaluate and verify the proposed application design. As this was a preliminary validation study, participants were recruited from the researchers' environments. The participant sample covers a diverse group of six females and four males, aged between seven and 63 years. All participants (and their legal guardians if underage) consented to participating in the validation experiment. Respectively, five out of 10 participants answered the questions "Are you familiar with smartwatches" and "Have you used assistant programs before (for example, for exercises)?" with "yes".

As one of the examples in the literature that could potentially benefit from JITAIs is compulsive hand-washing [24], we chose hand-washing interventions as the evaluation experiment. We show the flow of a single repetition of our experiment in Fig. 6. The study was designed so that participants were not informed of the nature of the intervention or the aim of the investigation, as this

[4] `https://github.com/RBurchard/JITAI-WizardAndroid`
[5] `https://github.com/RBurchard/JITAI-WizardControlStation`

might have biased their responses to the administered JITAIs. Thus, it is also a WoZ study. Before running the intervention experiment, we used our smartwatch application and the digital avatar "Odi" to administer instructions to the user, familiarizing them with the feedback modality and the mini-games before the actual experiment. Prior to administering the washing task, users were asked to dirty their hands by touching a harmless but noticeable substance, either moisturized coffee powder or blue finger paint. A previous study by Wahl et al used a similar method to artificially induce the urge to wash one's hands and to combat "repetition fatigue" [25].

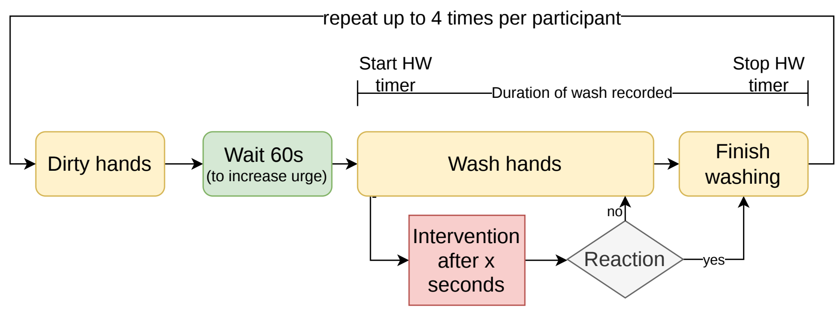


Fig. 6: The flow of a single repetition of our user study experiments. Participants repeated the steps, with different interventions being administered. We commence each repetition with a distraction task, for which the smartwatch displays some instructions. Participants then "dirty" their hands using either moisturized coffee powder or finger paint. They have to wait for 60 seconds to increase the hand-washing urge and are then asked to wash their hands. While the participant is washing their hands, an intervention is triggered by the experimenter, and the participant's reaction is recorded. After the run of the experiment, the participant is asked to fill in the questionnaire about the experience with the device.

After "dirtying" their hands, participants are asked to wait approximately 1 minute, thereby increasing their desire to wash their hands further. Then, they are instructed to wash their hands, and a timer is started by the experiment software to measure the hand-washing duration. During the hand-washing period, we administer an intervention after around 15 seconds, which is long enough for most participants to clean their hands reasonably well, but shorter than most participants would wash naturally, after dirtying their hands and enduring the waiting period. The intervention tries to stop the user from washing. We tested different configurations of interventions and alert types. As alert types, we employed vibration only, vibration with sound, and an escalating strategy: short vibration, then a short vibration and sound, and, as the last stage, a permanent vibration pattern and sound alerting. As mini-game distraction tasks, we tested “Simon Says”, “Stand Still”, “Lock Picking”, and interventions without mini-games. Although we could, e.g., relate the hand-washing durations and user reactions to the intervention and alert styles, we, for now, limit ourselves to gathering qualitative feedback, as the user study had a small sample size and we merely wanted to demonstrate feasibility.

The participant’s reaction to the intervention can either be to follow the instructions of the intervention or to ignore it and continue washing their hands. The reaction to the intervention and the total duration of the hand wash are logged in our application as the experiment progresses.

After the participant finishes washing their hands, they fill in the post-experiment questionnaire. We evaluate how the experiment and the interventions were perceived by the user, how helpful they found the smartwatch’s instructions, and how intrusive or frustrating the interventions appeared to them.

4 Results

4.1 Watch and Phone Applications

The applications for the smartwatch and for the smartphone lend themselves well to conducting experiments with JITAIs. As a side note, based on user feedback, we learned that they can also be used to provide general instructions to users. During our experiments, the connection between the smartphone and the smartwatch is stable, as we make use of the well-tested official Wear Data Layer API. Launching the interventions on the smartwatch from the background is also stable, and not a single intervention was dropped by the OS. While conducting experiments with the 10 participants, no crashes occurred. Because the WearOS application does not run in the background as a service, the battery life impact of the application is also minimal, as it is only ever launched when an intervention is sent through the Data Layer API. The watch can also collect physiological, health, and motion sensor data. If this is enabled, it consumes more energy, as all data is streamed to the smartphone and, optionally, to the desktop control panel.

4.2 User Study

The user study was conducted on a small sample size with 10 participants (6 female, 4 male, mean age 33.7, median age 29, min. age 7, max. age 63) and 29 repetitions of the intervention experiment. Below, we explain our general results with regard to the usability of our framework for delivering interventions. Fig. 7 shows questionnaire data collected from all participants. Responses indicated a generally high self-reported importance of hand hygiene (M = 7.90, SD = 2.13, median = 8.5, range = 5–10), measured on an 11-point scale from 0 ("not at all") to 10 ("very seriously"). Fig. 8 shows that the participants wash their hands for different durations, ranging from 10 s to 120 s. The effect of interventions on hand-washing durations could be measured with our applications in a future study.

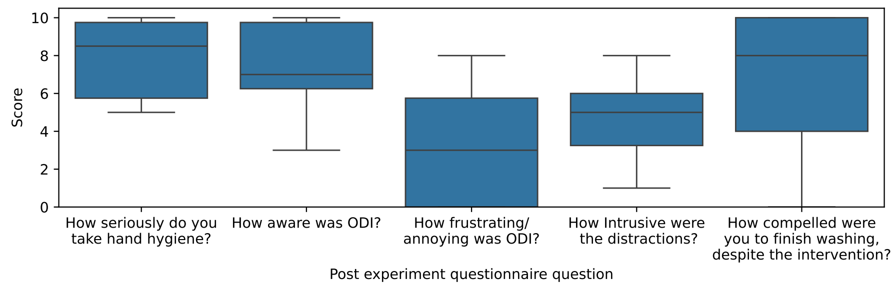


Fig. 7: Questionnaire questions answered by n=10 participants after conducting the experiments. Users were asked to rate on a Likert scale from 0 to 10, with 0 referring to a strong "not at all" and 10 referring to a strong "very".

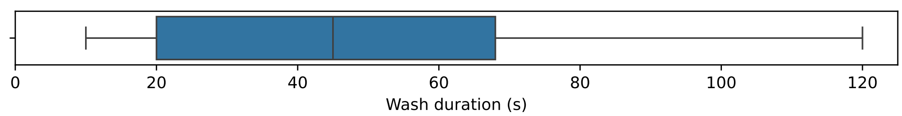


Fig. 8: Boxplot showing the distribution of wash durations among all repetitions.

Alert types We compare *vibration* to *vibration with sound*, and to the escalating alert. Out of 29 total interventions sent to the users, five were not registered by the user, all of them of the type *vibration only*. During the hand-washing activity, the watch's vibration was apparently not always noticed, as the user was moving and focused on washing hands. Adding the sound modality seems to help, as all other interventions were registered. Fig.9 shows how the users responded to each alert type.

Mini-games We evaluated three mini-games, which were generally understood and performed well by the participants. One issue appeared when a user had wet hands (from hand-washing), so that their inputs to the capacitive touch screen of the watch were misinterpreted.

In Fig. 10, we display the user's responses to interventions with and without attached mini-games. We only list interventions that were noticed by the user. Overall, the mini-games were enjoyed and completed often by the users, but none of them was effective in completely stopping the user from hand-washing.

User feedback As shown in Fig. 7, users generally responded positively to the idea of a digital assistant for adaptive interventions. As we did not reveal to them that "Odi" is manually controlled, many of them gave "Odi" high awareness ratings (median=7.5/10). Most users were not overly frustrated by "Odi"'s frequent interference. Users perceived the system as moderately distracting (me-

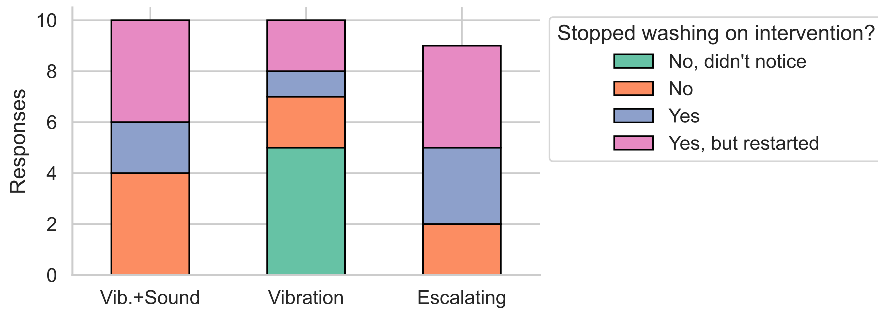


Fig. 9: Response to intervention during handwashing, by alert type. All alerts were noticed, unless otherwise indicated. The colors refer to the different sub-reactions: Some users ignored the notification, some continued washing after dealing with the notification and playing the mini-game, while others immediately stopped washing for good.

dian=5). Despite the interventions, almost all users felt compelled to finish washing, showing that the "dirtying" of their hands had the desired effect, and that interventions/distractions do usually not erase the desire to finish washing.

When asked whether "they could see themselves wearing a watch like this on a daily basis?", seven participants responded with "yes", three said "no". One of them said that they generally dislike wearing wristwatches, and another said they didn't feel any need for such an assistant, as they were not affected by OCD or negative habits. Other participants commented on liking the mini-games and the general idea of JITAIs, as the concept had been unknown to them.

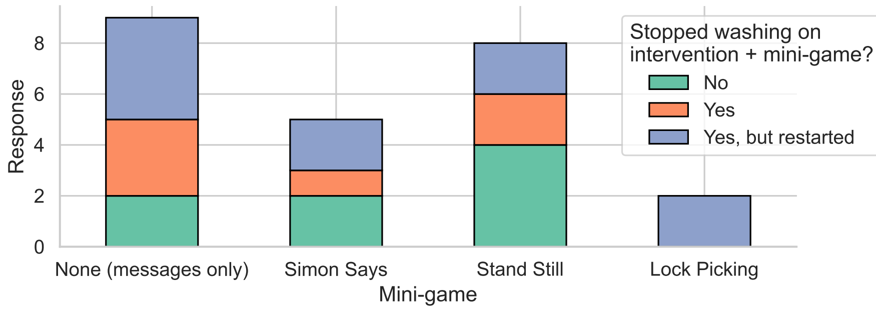


Fig. 10: Response to interventions with specific mini-games during handwashing, by type of mini-game. The plot shows only noticed interventions with/without prompts to play a mini-game. The colors refer to the different sub-reactions: Some users ignored the notification/mini-game, some continued washing after dealing with the notification and playing the mini-game, while others completed the mini-game and stopped washing, without recommencing to wash.

5 Code and Data Availability and Usage Notes

The complete source code of the Android and WearOS applications is free and openly available under the MIT License at GitHub[6]. To run an experiment using our toolkit, on the hardware side, an Android-based smartphone and a WearOS smartwatch are required.

The repository itself contains an extensive ReadMe.md file in which we explain all necessary steps to download, compile, and install the application(s) on the devices. Additionally, we include instructions for configuring and creating, saving, and loading experiment scripts directly on the smartphone via the app's GUI.

A secondary repository[7] contains the source code for the cross-platform desktop application. Its ReadMe.md file explains how to run the desktop control station.

Overall, by making available anonymized study data, experiment configurations, code, and instructions, we ensure a high level of reproducibility and a low entrance hurdle for reuse of our toolkit by other researchers.

6 Discussion

The qualitative results from our user study show that the applications are stable and adequate for conducting experiments similar to the one we described. More importantly, the validation demonstrates the feasibility of using JITAI-Wizard

[6] `https://github.com/RBurchard/JITAI-WizardAndroid`
[7] `https://github.com/RBurchard/JITAI-WizardControlStation`

as a rapid prototyping platform for wearable Just-in-Time Adaptive Interventions (JITAIs). Rather than evaluating a fully automated intervention system, our goal was to show that researchers can investigate intervention design questions before implementing complex sensing, activity recognition, and decision-making pipelines. As a contribution of this work, we make the applications freely available. By design, the smartphone app and the desktop application can load, modify, and save custom experiment scripts. Researchers from any domain can select from a multitude of intervention, message, or prompt parameters and use the smartwatch application to directly conduct their experiments, without any requirements for programming skills.

With our user study, we show that participants react well to the messages and instructions administered via the smartwatch, especially to sound-related notifications. The observation that vibration-only notifications were occasionally missed during hand-washing suggests that multi-modal notification strategies may be more suitable for activities involving continuous movement or reduced tactile perception. Overall, participants enjoyed the mini-games, and most could imagine using a system like ours in the future. While interventions were sometimes ignored, users in our example sometimes even continued to wash their hands *after* receiving the intervention and even after completing a mini-game. This indicates that engaging interaction alone is not necessarily sufficient to change ongoing behavior, particularly when users are strongly motivated to complete a task. At the same time, these findings further motivate the need for custom JITAI development by domain experts, who can use our system to develop and evaluate more effective alert patterns, intervention strategies, and prompt/distraction task combinations.

The users also gave high ratings for the clarity and perceived helpfulness of the interventions. The positive reactions suggest that users are generally receptive to receiving coaching and adaptive interventions through smartwatches. Although our evaluation focused on a hand-washing scenario, the same interaction principles could likely be transferred to other application domains beyond just-in-time adaptive interventions, such as assistance with manufacturing procedures, cooking recipes, rehabilitation exercises, or any other scenario where the user may benefit from instantaneous and context-aware feedback.

JITAI-Wizard enables researchers to investigate questions such as optimal intervention timing, notification modality, interruption tolerance, user receptivity, and intervention content before investing in automated sensing and decision-making pipelines. By allowing these design variables to be explored independently of recognition accuracy or previous dataset availability, the framework supports faster iteration and more informed development of fully automated JITAI systems. In this way, the Wizard-of-Oz methodology separates the evaluation of intervention design from the performance of automated context recognition, allowing each aspect to be studied independently and reducing the engineering effort required during early-stage research.

As explained in Section 5, we place particular emphasis on making available not only the source code, but also the documentation and setup instructions.

This allows other researchers to reproduce our experiments, adapt existing intervention protocols to their own studies, and extend the framework with additional functionality, ultimately lowering the barrier to reproducible research on wearable JITAIs.

Limitations While our application is built with customizability and extendability in mind, through our experiments, we were only able to verify its usefulness in our own and similar experiments. Other researchers might have other needs, and we did not run an analysis of possible needs or desired features for other researchers' projects. However, as the applications come with completely open source code and an open license, they can be extended to include additional features. Furthermore, although the framework supports a wide range of intervention types, notification modalities, mini-games, and parameter combinations, our validation study evaluated only a subset of these configurations. Consequently, the effectiveness and usability of many possible intervention designs remain to be investigated in future work.

Our application is a toolkit for other researchers to use, but its validation is also limited by the fact that we did not evaluate it with a third party, e.g., a team of psychologists. While we designed the applications to be user-friendly and easy to use, we cannot quantify the user experience other researchers might have.

Another limitation is the size of our user study, which we only scaled sufficiently large to obtain some preliminary findings on JITAI efficacy and to evaluate the stability of our applications. While the results hint at the users reacting to the administered JITAIs, a larger sample size would be needed to reliably measure or show a significant effect.

7 Conclusion

We present JITAI-Wizard, a free and open-source framework for designing and delivering wearable just-in-time adaptive interventions in Wizard-of-Oz studies. We describe the system architecture, information flow, intervention capabilities, and the design choices that enable researchers to rapidly prototype and evaluate wearable intervention concepts without requiring a fully automated sensing and decision-making pipeline. A key contribution of this work is the explicit decoupling of intervention design from automated sensing and recognition, enabling early-stage JITAI research without requiring full context-awareness systems. JITAI-Wizard is designed to be used without programming expertise, allowing experiment scripts to be created, modified, shared, and executed directly through graphical smartphone and desktop interfaces.

The framework supports a wide range of intervention modalities, including informational prompts, feedback messages, timers, and interactive mini-games, making it suitable for diverse research scenarios. Through a validation study involving induced hand-washing behavior, we demonstrate that the platform can reliably deliver interventions in real time and collect detailed response data.

Participants generally perceived the smartwatch-delivered interventions as clear, noticeable, and helpful, although results remain preliminary and should be interpreted in light of the limited sample size and evaluated intervention configurations.

By releasing the complete source code, documentation, and experimental artifacts under an open-source license, we aim to lower the barrier to entry for researchers interested in JITAIs, digital health interventions, and wearable user studies. Future work includes larger-scale evaluations, integration with automated activity recognition and context-detection pipelines, and exploration of additional application domains beyond health behavior change. The study also highlights that the notification alert modality can substantially affect user perception and responsiveness, indicating that multi-modal and context-sensitive intervention strategies remain an important direction for future research. We hope that JITAI-Wizard will facilitate faster iteration, improved reproducibility, and broader adoption of wearable intervention research.

Acknowledgments. We thank all participants of our evaluation study.

Disclosure of Interests. The authors declare that they have no competing interests.

References

1. Babu, M., Lautman, Z., Lin, X., Sobota, M.H., Snyder, M.P.: Wearable Devices: Implications for Precision Medicine and the Future of Health Care. Annual Review of Medicine **75**(1), 401–415 (Jan 2024). `https://doi.org/10.1146/annurev-med-052422-020437`
2. Carter, S., Mankoff, J., Heer, J.: Momento: Support for situated ubicomp experimentation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 125–134. ACM, San Jose California USA (Apr 2007). `https://doi.org/10.1145/1240624.1240644`
3. Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of Oz studies — why and how. Knowledge-Based Systems **6**(4), 258–266 (Dec 1993). `https://doi.org/10.1016/0950-7051(93)90017-N`
4. Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J., Gandy, M.: Wizard of Oz support throughout an iterative design process. IEEE Pervasive Computing **4**(4), 18–26 (Oct 2005). `https://doi.org/10.1109/MPRV.2005.93`
5. Ferreira, D., Kostakos, V., Dey, A.K.: AWARE: Mobile Context Instrumentation Framework. Frontiers in ICT **2** (Apr 2015). `https://doi.org/10.3389/fict.2015.00006`
6. Fischer-Janzen, A., Wendt, T.M., Görlich, D., Van Laerhoven, K.: Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices. J. Hum.-Robot Interact. **15**(3), 56:1–56:31 (Apr 2026). `https://doi.org/10.1145/3796524`
7. Gustafson, D.H., McTavish, F.M., Chih, M.Y., Atwood, A.K., Johnson, R.A., Boyle, M.G., Levy, M.S., Driscoll, H., Chisholm, S.M., Dillenburg, L., Isham, A., Shah, D.: A smartphone application to support recovery from alcoholism: A randomized clinical trial. JAMA psychiatry **71**(5), 566–572 (May 2014). `https://doi.org/10.1001/jamapsychiatry.2013.4642`

8. Intille, S., Haynes, C., Maniar, D., Ponnada, A., Manjourides, J.: μEMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 1124–1128. UbiComp '16, Association for Computing Machinery, New York, NY, USA (Sep 2016). https://doi.org/10.1145/2971648.2971717
9. Kelley, J.F.J.: Wizard of Oz (WoZ): A yellow brick journey. Journal of Usability Studies **13**(3), 119–124 (May 2018)
10. Klasnja, P., Consolvo, S., Pratt, W.: How to evaluate technologies for health behavior change in HCI research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 3063–3072. CHI '11, Association for Computing Machinery, New York, NY, USA (May 2011). https://doi.org/10.1145/1978942.1979396
11. Klasnja, P., Smith, S., Seewald, N.J., Lee, A., Hall, K., Luers, B., Hekler, E.B., Murphy, S.A.: Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine **53**(6), 573–582 (May 2019). https://doi.org/10.1093/abm/kay067
12. Künzler, F., Mishra, V., Kramer, J.N., Kotz, D., Fleisch, E., Kowatsch, T.: Exploring the State-of-Receptivity for mHealth Interventions. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **3**(4), 140:1–140:27 (Sep 2020). https://doi.org/10.1145/3369805
13. Li, I., Dey, A., Forlizzi, J.: A stage-based model of personal informatics systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 557–566. CHI '10, Association for Computing Machinery, New York, NY, USA (Apr 2010). https://doi.org/10.1145/1753326.1753409
14. Mast, D., Roidl, A., Jylha, A.: Wizard of Oz Prototyping for Interactive Spatial Augmented Reality in HCI Education: Experiences with Rapid Prototyping for Interactive Spatial Augmented Reality. In: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. pp. 1–10. CHI EA '23, Association for Computing Machinery, New York, NY, USA (Apr 2023). https://doi.org/10.1145/3544549.3573861
15. McNamee, P., Murray, E., Kelly, M.P., Bojke, L., Chilcott, J., Fischer, A., West, R., Yardley, L.: Designing and Undertaking a Health Economics Study of Digital Health Interventions. American Journal of Preventive Medicine **51**(5), 852–860 (Nov 2016). https://doi.org/10.1016/j.amepre.2016.05.007
16. MISHRA, VARUN., KÜNZLER, FLORIAN., KRAMER, J.N., FLEISCH, ELGAR., KOWATSCH, TOBIAS., KOTZ, DAVID.: Detecting Receptivity for mHealth Interventions in the Natural Environment. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies **5**(2), 74 (Jun 2021). https://doi.org/10.1145/3463492
17. Nahum-Shani, I., Murphy, S.A.: Just-in-Time Adaptive Interventions: Where Are We Now and What Is Next? Annual Review of Psychology 2026 pp. 679–703 (2026). https://doi.org/10.1146/annurev-psych-121024044244
18. Nahum-Shani, I., Smith, S.N., Spring, B.J., Collins, L.M., Witkiewitz, K., Tewari, A., Murphy, S.A.: Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine **52**(6), 446–462 (May 2018). https://doi.org/10.1007/s12160-016-9830-8

19. Nezami, B.T., Valle, C.G., Wasser, H.M., Hurley, L., Hatley, K.E., Tate, D.F.: Optimizing a mobile just-in-time adaptive intervention (JITAI) for weight loss in young adults: Rationale and design of the AGILE factorial randomized trial. Contemporary Clinical Trials **150**, 107808 (Mar 2025). https://doi.org/10.1016/j.cct.2025.107808
20. Riek, L.D.: Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. Journal of Human-Robot Interaction **1**(1), 119–136 (Jul 2012). https://doi.org/10.5898/JHRI.1.1.Riek
21. Sarker, H., Sharmin, M., Ali, A.A., Rahman, M.M., Bari, R., Hossain, S.M., Kumar, S.: Assessing the availability of users to engage in just-in-time intervention in the natural environment. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 909–920. UbiComp '14, Association for Computing Machinery, New York, NY, USA (Sep 2014). https://doi.org/10.1145/2632048.2636082
22. Schlögl, S., Doherty, G., Luz, S.: Wizard of Oz Experimentation for Language Technology Applications: Challenges and Tools. Interacting with Computers **27**(6), 592–615 (Nov 2015). https://doi.org/10.1093/iwc/iwu016
23. Tangmunarunkit, H., Hsieh, C.K., Longstaff, B., Nolen, S., Jenkins, J., Ketcham, C., Selsky, J., Alquaddoomi, F., George, D., Kang, J., Khalapyan, Z., Ooms, J., Ramanathan, N., Estrin, D.: Ohmage: A General and Extensible End-to-End Participatory Sensing Platform. ACM Transactions on Intelligent Systems and Technology (TIST) **6**(3), 38:1–38:21 (Apr 2015). https://doi.org/10.1145/2717318
24. Wahl, K., Scholl, P.M., Miché, M., Wirth, S., Burchard, R., Lieb, R.: Real-time detection of obsessive-compulsive hand washing with wearables: Research procedure, usefulness and discriminative performance. Journal of Obsessive-Compulsive and Related Disorders **39**, 100845 (Oct 2023). https://doi.org/10.1016/j.jocrd.2023.100845
25. Wahl, K., Scholl, P.M., Wirth, S., Miché, M., Häni, J., Schülin, P., Lieb, R.: On the automatic detection of enacted compulsive hand washing using commercially available wearable devices. Computers in Biology and Medicine **143**, 105280 (Apr 2022). https://doi.org/10.1016/j.compbiomed.2022.105280
26. Wang, L., Miller, L.C.: Just-in-the-Moment Adaptive Interventions (JITAI): A Meta-Analytical Review. Health Communication **35**(12), 1531–1544 (Oct 2020). https://doi.org/10.1080/10410236.2019.1652388
27. Wang, R., Zhang, K., Wang, Y., Brown, K., Zhao, Y.: "It was Mentally Painful to Try and Stop": Design Opportunities for Just-in-Time Interventions for People with Obsessive-Compulsive Disorder in the Real World. In: Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 1–18. ACM, Denver Colorado USA (Oct 2025). https://doi.org/10.1145/3663547.3746394
28. Xiong, H., Huang, Y., Barnes, L.E., Gerber, M.S.: Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 415–426. UbiComp '16, Association for Computing Machinery, New York, NY, USA (Sep 2016). https://doi.org/10.1145/2971648.2971711
29. Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., Alshurafa, N.: Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. Sensors **22**(4), 1476 (Jan 2022). https://doi.org/10.3390/s22041476