Wearables in the Wet Lab:
A Laboratory System for Capturing and Guiding Experiments

Philipp M. Scholl
Embedded Systems, Albert-Ludwigs-Universität Freiburg
pscholl@ese.uni-freiburg.de

Matthias Wille
Unit Human Factors, Ergonomics Federal Institute for Occupational Safety and Health Dortmund, Germany
dr.matthias.wille@gmail.com

Kristof Van Laerhoven
Embedded Systems, Albert-Ludwigs-Universität Freiburg
kristof@ese.uni-freiburg.de

ABSTRACT

Wet Laboratories are highly dynamic, shared environments full of tubes, racks, compounds, and dedicated machinery. The recording of experiments, despite the fact that several ubiquitous computing systems have been suggested in the past decades, still relies predominantly on hand-written notes. Similarly, the information retrieval capabilities inside a laboratory are limited to traditional computing interfaces, which due to safety regulations are sometimes not usable at all. In this paper, Google Glass is combined with a wrist-worn gesture sensor to support Wetlab experimenters. Taking "in-situ" documentation while an experiment is performed, as well as contextualizing the protocol at hand can be implemented on top of the proposed system. After an analysis of current practices and needs through a series of explorative deployments in wet labs, we motivate the need for a wearable hands-free system, and introduce our specific design to guide experimenters. Finally, using a study with 22 participants evaluating the system on a benchmark DNA extraction experiment, we explore the use of gesture recognition for enabling the system to track where the user might be in the experiment.

Author Keywords

Life Science; Hands-free Documentation; Google Glass; Wrist-Worn Inertial Sensors; Wet Lab; Activity Recognition

ACM Classification Keywords

H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

INTRODUCTION

Much of the experimental research in life sciences is performed in wet labs. These laboratories are highly dynamic
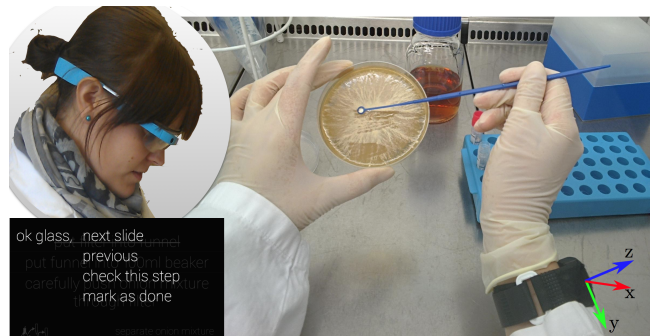

Figure 1. A hands-free wearable system combines Google's Glass (top-left) with a wrist-worn accelerometer logger (lower-right) to allow capturing and reviewing of experiments in a wet laboratory environment.

in nature and contain benches packed with equipment such as tubes, racks, compounds, specialized machinery and computers. Chemicals, organisms, liquids and other agents are experimented on. The experiments performed in such labs can take a few minutes up to several days, and need to suspended often for organism to grow, or until a sufficient amount of agents are incubated. To provide valid results, experiments are required to be reproducible, which is why experimenters rely on checklist-like documentation notes taken during or prior to their procedures. Off-the-shelf Kits, which implement common procedures, provide detailed usage instructions. Other details of procedures are either established from scratch, adapted from similar experiments, incrementally refined as experiments get repeated or executed as close as possible in an experiment reproduction cycle. The observations of the scientist and details of the procedure are recorded in so called *protocols*.

Maintaining these lab protocols often ends up as a heavy burden for the experimenter, resulting subsequently in protocols that are routinely written "offline", after the experiment has been performed. Individual steps and details are mostly reconstructed from memory, if written down at all. One reason for this, is the large effort in putting down the experimentation equipment (e.g., pipettes, flasks, gloves, or containers) and moving to a different bench to take notes or retrieve information from a PC. Another more profound reason is the risk of contamination - laboratories are classified into four bio-safety levels with increasing precautions to contain harmful agents. Easily contaminated and absorbent materials

Figure 2. Examples of the user interface as presented to the participants while performing the study. Information cards are switched backwards and forwards, the current step can be marked as complete using speech commands. The second and fourth images are examples of checked /marked steps.

like paper, lab notebooks, cameras and other utilities need to be sterilisable. Wrapping these into easily de-contaminated materials (e.g. shrink-wrapping), not allowing them to be taken out of the laboratory or requiring them to be used only with protective garments are common routines. Even when working with non-harmful agents, contamination can be an issue when materials are shared or compounds transferred between experiments via contaminated tools. Such contaminations can easily invalidate experimental results. Therefore, the taken safety pre-cautions serve a double purpose: (1) to protect the experimenter from the organism and (2) the organism under study from unwanted contaminations from the outside. These neccesary pre-cautions however hinder the on-the-spot documentation retrieval and generation abilities of scientists in the wet lab.

We follow in this paper a similar approach to that of a set of previous studies from ubiquitous computing research, which posit that assisting the experimenters with an interactive and augmented lab environment has strong potential to increase their efficiency, to decrease the probability of errors, and to allow for a comfortable recording of experimental parameters in biology or chemistry experiments. We present to this end a *wearable system* providing features such as a head-up display, point-of-view video recordings, voice recognition and hand gesture recognition which allows inconspicuous interaction with the laboratory. It also provides the opportunity to assist individual experimenters by equipping themselves rather than changing the environment for all its users.

More specifically, we present in this paper a fully wearable system consisting of Google Glass and a wrist-worn accelerometer, to aid in documenting the experiment steps in a wet laboratory environment. Our core assumption is that the usefulness of continuous streams of video, audio or environmental sensor data is increased, when its searchable by human-understandable cues or can be retrieved automatically by current context. In this paper we ask wether wrist accelerations can provide such cues. The following three contributions are made:

- A case study of a microbiology laboratory is introduced, which studies the challenges a wearable system needs to overcome, via a series of observations of point-of-view video recordings in different lab environments.

- We present a hands-free system that enables wet lab scientists to record their work, or look up, navigate, and mark steps of an existing protocols through a head up display.

- The feasibility of detecting common gestures with 22 novice users from wrist acceleration readings in a controlled laboratory study (DNA extraction) is shown. Fur-

thermore reconstructing large parts of the experimental protocol based on key getures (such as mixing, stirring, or pouring) by a Hidden Markov Model is demonstrated.

In the remainder of this paper, the system with its individual components is described and a deployment of the system for capturing experiments in several types of wet laboratories is discussed. We then provide a study on what is achievable with gesture recognition algorithms alone as a capturing approach, followed by a discussion of the results with relation to future directions that this approach can take. First, we compare our wearable system to other systems for wet laboratories that have been proposed in the past years, and more generally to wearable assistant systems for other domains.

RELATED WORK

The notebook, in which experimenters record their thoughts, results, and plans serves as one of the core parts during a life scientists' research. As such it has been the target of many research efforts, as well as targeted by commercial endeavours. Subsequently, electronic laboratory notebooks (ELNs) are often sought to replace or enhance their pen and paper counterparts. Even when offering clear advantages, like *searchability* of records, direct addition of *multimedia*, *eased collaboration* [1–3], *edit-ability* [4] and *capturing of instrument measurements* [5], their usage is still limited. Since these notebooks are used for tracing and claiming inventions, as well as checking conformance to established protocols, liability and legal issues arise with their usage. For electronic records, only a limited legal framework is available [6, 7], limiting possible usage scenarios only to some laboratories.

Pen and paper solutions are also preferred because of the flexibility and freedom over visual structure they provide [8]. This has lead to several efforts which try to combine physical and electronic notes. The *a-book* [9] combines a tablet and PDA to capture paper notebook writing. Based on a fiducial marker, entries could be augmented with additional media and easily shared. A system to support biologists in the field was presented in the *ButterflyNet* [10] project. Handwritten notes are captured (with an Anoto pen) and combined with visual and audio information for later access. This allowed the biologist to capture information in the field, and augment it with other sensory clues - a task that previously had to be done manually. The *Prism* [1] project reports on a study of biologists' work practices and presents a hybrid system using hand-written notes as well as digital content to capture, visualise and interact with activity streams in the laboratory. Forcing too much structure has been found to be too inflexible. An open design based on linking and searching information bits was adopted, similar to the MyLifeBits design [11] but specific to the experimenters workflow.

Other type of systems are focussed on providing a more formal specification of single experimental workflows, on which user interfaces are built. These systems also augment the lab itself to provide these user interfaces. The *LabScape* [12] project was an early investigation in a ubiquitous computing platform to help scientists and students to access and capture information in the laboratory. It uses interactive flowchart diagrams to visualize and annotate ongoing procedures that are accessed via a touch-tablet, barcode scanner, RFID tags [13], numeric keypad and wireless keyboard.

During the *Combechem* project the idea of a digitized flowchart was enhanced to a Semantic Smart Laboratory [14], a system for supporting chemistry experiments focusing on providing a flexible ontology for describing experiments and storing them for later retrieval. A formal definition of chemical experiments is presented as part of the *Labtrove* project [15] as well. The *eLabBench* [16] and *Biotisch* [17] takes the integration in the laboratory further by replacing the traditional workbench with a tabletop system that presents information on the bench's surface, also allowing interaction, sensing of augmented objects (e.g, racks of test tubes) and taking pictures of the whole setup with an overhead camera. The gathered digital information is stored in a wiki-like notebook for later retrieval.

In contrast, the system that is proposed in this paper focuses on the largely unexplored area of supporting and augmenting such laboratory tasks by means of a lightweight, exclusively wearable system. The setup requires little to no interference with the laboratory environment and its inventory, and offers hands-free operation. We argue that this approach of augmenting the *researchers* instead of the laboratory, has many advantages, not in the least the fact that every user in existing laboratories can still opt to keep on documenting their experiments with the traditional methods alone [18].

As has been shown by some of the above research, a formal workflow tends to be valued by experimenters and can be exploited as a structure for information capture as well. We investigate in particular whether a wrist-worn accelerometer unit can be used to capture such pre-defined structure, for instance in order to index associated video and audio recordings. This concept of automated journal building has been introduced previously, where activities were extracted in an unsupervised fashion [19]. In the case of the life science laboratory, these workflows tend to be frequently predefined, and actions can be extracted from available textual descriptions, leading to a semi-supervised approach.

The concepts of wearable workflow monitoring, documentation access, and assistance have been thoroughly explored by research in other domains [20], especially in maintenance [21], manufacturing [22] and inspection tasks [23]. Key components of the envisioned system in this paper were also inspired by the Remembrance Agent (Remem) project at the MIT more than a decade ago. The proposed system is however focused on the specific scenario of documenting and assisting wet laboratory tasks, where contextual knowledge is well-defined, and tasks and information have additional constraints that the lab environment poses.

Figure 3. Four selected steps during the mock-up DNA Extraction study. The participants are wearing a wrist accelerometer and Google's Glass. The latter guides the participant through the experiment.

SYSTEM DESIGN

For the work presented in this paper, we implemented the basics of a voice-interacted task guidance and logging system. (cf. Figure 2) The system is able to display workflow steps, and navigate or mark steps as done through voice interaction. It also includes the ability to capture audio and video from Glass , and interactions with Glass in the background. Additionally gestures made by the dominant hand from a low-power inertial sensing unit (cf. Figure 3) are recorded. The data of both systems is merged offline, as this implementation was only used for this particular experiment. We will present both system components in this section and detail their implementations.

The Wrist-Worn Unit

A wrist watch-like 3D acceleration sensing unit is used to capture motion sequences from the experimenter's dominant hand [24]. It consists of a low-power micro-controller connected to an ADXL345 3D-accelerometer. Acceleration data is logged in a compressed format to an on-board microSD-card and the included 180mAh-battery can keep the system running for at least 2 weeks continuously. The data from the wrist-worn unit is in the system's current implementation retrieved after the experiment are completed via USB, and analyzed offline for gestures that can be correlated to certain experiment steps. The device is reprogrammable via USB and protected by a plastic enclosure, visible in the lower-right section of Figure 1. The sampling rate was set to 50Hz, and collected in $\pm4g$ range, for the studies described in the deployment and experiment sections hereafter.

Google's Glass

Glass was used to capture experiments via its side-mounted video camera and microphone. Additionally, an application was designed to guide experimenters through predetermined experiment protocols. For the latter, an example user interface presented to participants can be seen in Figure 2 and shows some of the steps of the DNA extraction protocol that was used during this paper's study. Single steps

Figure 4. The deployment of the system combining Glass and the wrist-worn accelerometer, while continuously recording researchers at work in a wet lab. The environment is often simultaneously used by large groups of researchers or students and is equipped with a multitude of shared instruments and special safety zones, making it challenging to augment the environment (top half). Hands-free recording is a strong advantage: Often, experiments require gloves for minimizing contamination risks; Wet labs furthermore contain a large variety of compounds, instruments and lab equipment that require both hands to be used (bottom half). All these photos are unaltered frames taken from continuous Glass video recordings.

in the protocol are shown in a timeline, that could be navigated to the left (for past steps) and to the right (for future steps). The wearer has the choice to navigate the protocol using either the swipe gestures on the touchpad[1] on the Glass, or by voice commands. The subset of voice commands chosen for this were "previous", "next slide", "check this step" and "mark as done". These commands were chosen by experimenting with their recognition rates, trying to increase their phonetic dissimilarity for multiple speakers. By saying "check this step", "mark as done" or by tapping the touchpad, the user can let the system know that the step was performed, which is visualized by striking the current item through. The "ok, glass" guard phrase, which is usually required for Glass apps, was removed to minimize interaction time. If this item was the last step on the slide, the system automatically displays the next step. The application was implemented using the Android Framework, and works on Google's Glass as well as on Tablets and Smartphones.

The guiding part of our system is designed to be easily adaptable to different protocols. For this, we decided to use a document-driven approach, in which a human-readable and machine-parseable document contains the steps of a procedural protocol. These steps are written down in Markdown [25] documents. This allows experimenters to modify and present these workflow steps on different personal computing accessories (like PCs, laptops, Smartphones, and Wear-

ables) without much implementation effort. It furthermore allows for linking documents, and referencing additional media files. However, the major reason for using a document-driven design is that modification of the protocol can be captured easily in a distributed fashion. Only the transformations of the document need to be transported to a central repository. The protocols can then be shared cross-device and can be scoped on a per-user basis. This provides the means to share, collaborate on and synchronize the experimental protocol.

The interaction design of a wearable system for recording biology experiments is another important aspect. It should be possible to operate in a hands-free manner and stay out of the user's way most of the time (after all, the experimenter is interested in the outcome of his/her experiment and not recording his efforts in all detail). Explicit interaction can, for example, be achieved through voice recognition, or touch-free action/gesture/motion recognition. Since such an explicit interaction requires the user's full attention it should be used only when necessary. This is the reason why we also take into account any possible implicit interactions, through continuously monitoring the user's wrist movement to infer his or her context to provide additional information. Such contextual information can also be used to provide cues from large bodies of recorded data, similar to life-logging applications [11]. For example, imagine an experimenter video-recording all of his experiments and afterwards being able to efficiently jump to all sequences where something was pipet-

---

[1]The touchpad was enabled in this study, to avoid problems with voice recognition for non-native english speakers.

ted or to particular steps in his protocols that have been detected via his wrist movements.

DEPLOYMENT AS A RECORDING TOOL

The system as described in the previous section was deployed for recording the users at work in three different wet laboratory environments: an academic training laboratory, an academic research laboratory, and a commercial laboratory. Four recording sessions of up to a full day were held. In each recording, up to five microbiology researchers or students were wearing the system simultaneously, working within groups of up to four persons. Figure 4 shows some examples of the video footage taken with Glass. All users were at all times aware of the recordings (Glass showing the current recording) and were encouraged to discuss their experimentation steps and methods, and to provide feedback of the system and its possible advantages and disadvantages. The sound recording's quality of Glass was good enough in all environments to understand both the user and the people in the immediate proximity. After the system was handed out and activated, we did not remain present in the laboratories, and examined the contents of the videos afterwards.

In general, the acceptance of wearing the system was high, and even in the teaching laboratory, where approximately 20 fellow students were working in the same immediate environment of the user, nobody expressed concerns about the possibility of them being recorded. The latter observation might be due to the fact that experimenters do frequently take their personal cameras with them to photograph or record important experiment results. Apart from a few remarks made towards the end of the teaching sessions (which lasted over three hours each), we did not note any big signs of discomfort in wearing the system: Glass was at two occasions taken off to concentrate on using a microscope, and once to demonstrate it to a fellow user. One of the wrist-worn accelerometer sensors did not record consistent data as it was not strapped on tight enough and had rotated along the wrist during the course of the experiments. The video quality of Glass (at 720p) tends to be good enough to be able to read most compound labels and handwritten notes.

Several findings that emerged from the video footage are especially noteworthy: (1) Even in laboratories with a lower safety clearance which implies minimal contamination risks and therefore does not require gloves, the option of taking pictures or videos hands-free is a strong advantage. On many occasions, users required both hands simultaneously to handle instruments and the fact that Glass was able to record from a first-person perspective was at several occasions mentioned as a great feature. (2) The use of pen and paper notes is still largely preferred as a primary capturing system. Partly, this is due to its flexibility, but the videos also made clear that ad-hoc written notes, labels, instructions and lab books are truly ubiquitous in the wet lab. A digital system for providing assistance in these surroundings, apart perhaps from some tightly-regulated laboratories, has more chance of adoption when introduced as a complementary technology. (3) The ability of following what is recorded by Glass in the peripheral display was at multiple times used to guide the capturing of the video. When looking through a microscope,

	task	actions
1	**solvent** combine 50ml lukewarm water, 1/2 teaspoon salt and 3ml dishsoap in 200ml beaker and stir	pouring, transfer, pipetting, stirring
2	**cutting** peel and cut onion/tomato	peeling, cutting
3	**mixing** mix into 200ml beaker, add 1ml detergent, and stir	pouring, pipetting, stirring
4	**waterbath** put 200ml beaker into hot waterbath for 10mins	
5	**waterbath** put 200ml beaker into cold waterbath for 5mins	
6	**pestling** pestle mixture	pestling
7	**filtrate** put filter into funnel, funnel into 100ml beaker, push mixture through filter	
8	**pouring** pour 1.5ml of mixture into test tube, mix in 5ml freezing ethanol	pouring, pouring
9	**detection** carefully invert test tube multiple times	inverting

Table 1. The (shortened) DNA extraction protocol as shown to participants on Google Glass. The protocol was interleaved for both an onion and tomato, creating 18 steps in total. Gestures used to detect each step in the protocol are show in the right column.

for instance, several users used their Glass' display to make a recording through the eyepiece. Instead of taking immediate notes on compound quantities, e.g., to record how many millilitres of a solution were obtained, users would hover it closely to Glass' camera. Both during and after the deployments, many of the users were interested in using the system for subsequent times and expressed that they could envision continuing using it for capturing their experiment work.

The idea for this deployment was to get first insight into the working environment of experimental biologists, and also get a first idea of the usability of using Glass for recording only. To this end, we elicited challenges from several point-of-view video recordings from several laboratories. One specific challenge, that was also mentionend by participants, is the navigation of such continuous recordings. In the remainder of this paper we will look at the possibility of using typical hand gestures to detect the actions conducted during an experiment. These actions, extracted from a procotol's description, will then be used to extract time-codes than can index such continuous recordings - serving as the cues to navigate these recordings and providing guidance just-in-time.

EVALUATIONS IN EXPERIMENT GUIDANCE

We chose DNA extraction, a common entry-level laboratory experiment, for testing the feasibility of our system for both guiding novice users through an experiment and recognizing single activities in an experimental procedure using the wrist-worn sensor. The experiment is guided by a protocol, visualized as a textual step-by-step guide on the Google Glass display. It contains a sequence of two extractions of DNA, first that of an onion and second the DNA of a tomato, with the basic procedure being identical for both. This way all actions are repeated at least once with different material.
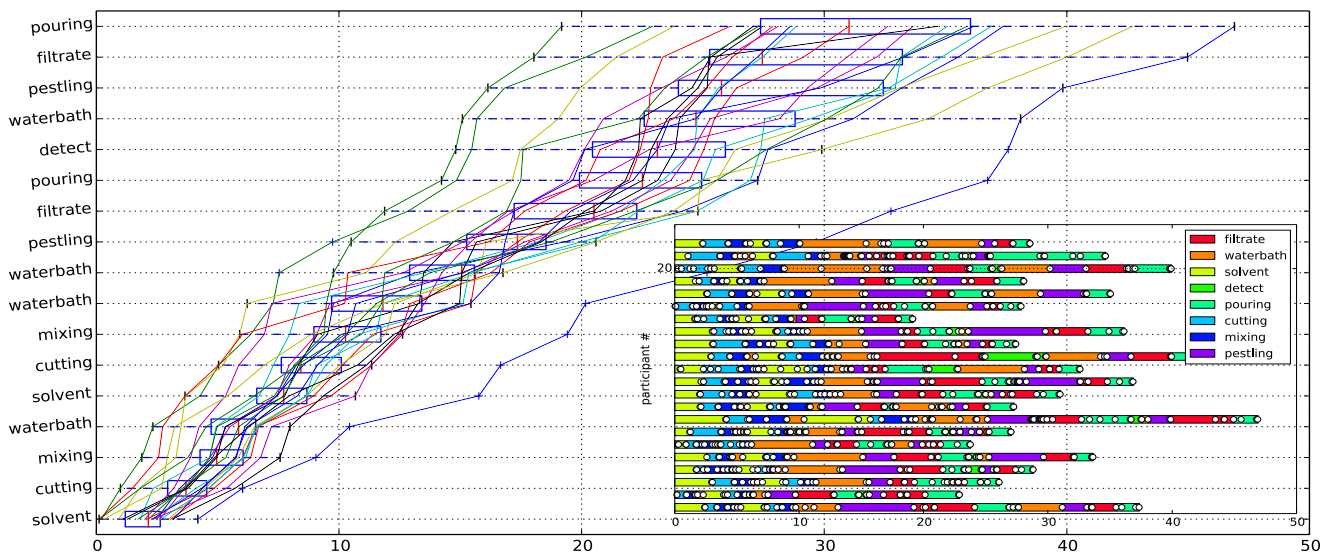
Figure 5. The mean duration of protocol steps per participants. The figure in the background shows the per-step mean duration across all participants (box-and-whisker) and the individual flow for each participant (lines). The figure in the bottom right shows the mean duration of each step per participant, color-coded to individual steps in the DNA Extraction protocol. Note that each step is repeated twice and interleaved during the experiment, once for the onion and once for the tomato. Markers on this figure show when an actual user interaction happened (when marking a step as done for example). The x-axis on both figures is the time taken in minutes.

The two experimental procedures were interleaved to save time, and each participant had to complete 18 steps in total, containing 9 different protocol steps to identify (shown in Table 1) and executed two times per participant. Each protocol step was described and displayed when activating the screen of Google Glass (by tilting the head up or tapping Glass' swipe area). Participants were instructed to primarily interact with Glass by speech and to move through the experiment protocol by marking each single step as done. In case the speech recognition would prove to be impractical, touch interaction (tapping) and swiping back and forth for moving between steps was kept as a backup option. The time for which a step was active on Glass' display was recorded in a log file, and is assumed to be the time it took to go through the displayed instructions. Figure 3 shows the experimental setting before and during the experiment, as recorded from a camera that was mounted at the ceiling.

In total, 22 participants took part in our experiments. Participants were recruited via advertisements in a local newspaper, representing persons with no prior experience in biology experiments and no affiliation to our research. The experiment was run in an environment for performing user studies, with video cameras recording the working bench. Before the task started, the participants were introduced to the functionality of the Google Glass and how they should use it during the DNA Extraction, what the different ingredients are e.g. ethanol or detergent, and where to find them. The participants were then fitted with Glass and the wrist-worn sensor, and asked to follow each displayed step and mark them as done as soon as they are completed. Following an experiment protocol took the participants between 18 and 45 minutes. A successful experiment would result in the DNA becoming visible as a set of small stripes in a test tube, although for our evaluation it did not matter whether the DNA extraction was successful in the end.

The choice of novice users instead of professional experimenters might be suprising. The goal of this study however, was to show that hand motion sequences can be used to detect protocol steps and actions. Especially to create a dataset that can serve as a benchmark for different detection algorithms. It is therefore important to have a high variabilty in executing different actions, as this is the case also for professional experimenters, i.e. everybody has their own styles. More execution variabilty also means a harder challenge for the detection system, so if it works for untrained personell it will most likely work for trained personell as well. Also, since participants were non-trained, they also adhere stricter to the protocol, a professional in turn might take shortcuts in the experiment since he is aware of the overall goal and working of the experiment. This would create datasets with different execution that are harder to compare. The recorded dataset can therefore be used as a baseline benchmark for recognizing actions that are related to those in a wet laboratory - in a systematic manner.

The goals of this study were threefold: First, we looked into the specific interactions with the guidance applications to see different usage patterns of participants. Second, we want to show how well actions (defined as repeated motion sequences) can be detected through wrist movement measured by an accelerometer - or put differently, how discriminative the measured data is when applied to typical workbench activities. Third, we wanted to know whether the detectable motion sequences (or action sets) correlate with specific protocol steps. Our interest is first and foremost in knowing how well a system could detect the combination of recording interactions, while guiding people through a wet laboratory experiment, and wrist movements. This could be used to automatically detect steps in wet laboratory workflows.
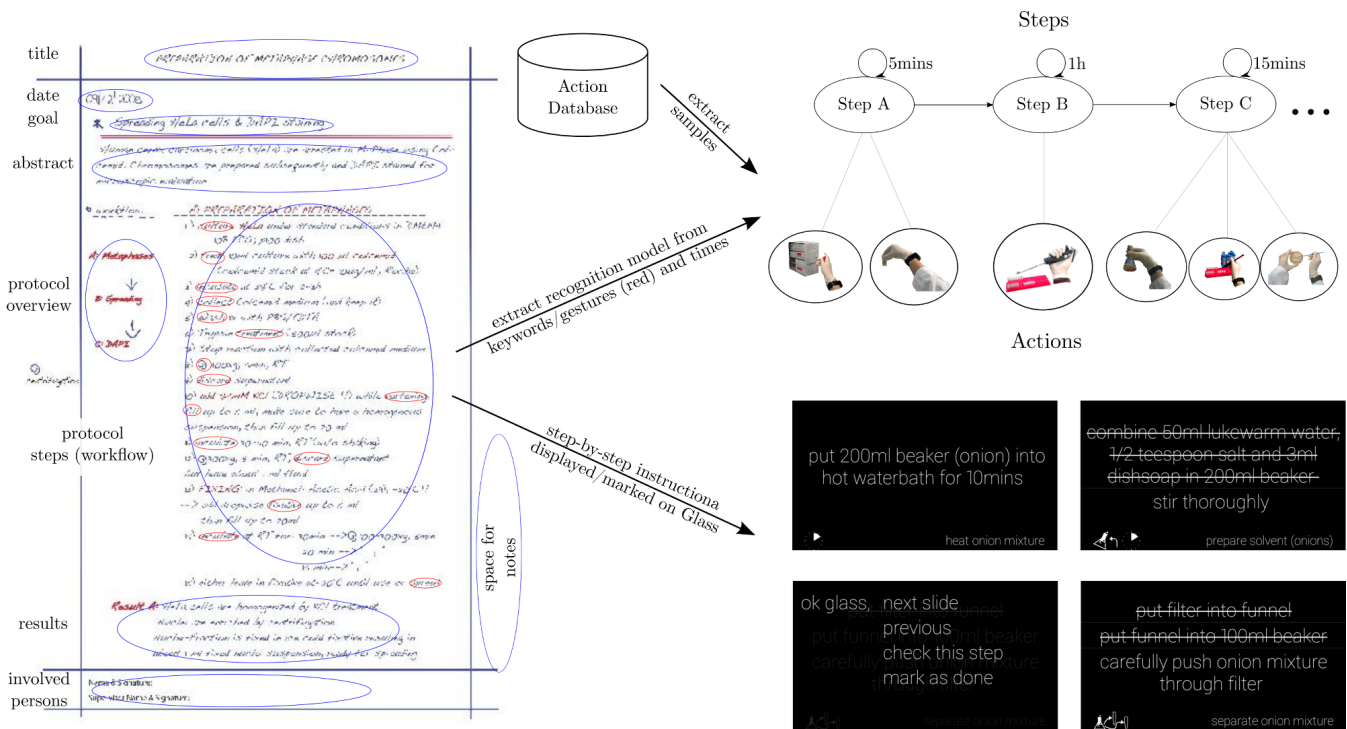
Figure 6. A (typical) page in laboratory notebook and the possibly extracted recognition and guidance system. An action database contains recordings of wrist motion samples. This database will be used for action detection, which in turn serves as the observations of a Hidden Markov Model, which contains each step in the protocol as a hidden state. Each protocol steps contains the to be detected actions. Time is implicitly encoded via the number of made observations. Each protocol step can be displayed on Glass for guidance.

User Interaction with the Glass Application

During the whole experiment, every interaction (both voice commands and swipes) with the Google Glass application was logged for later investigation. As this was done in the background, participants were unaware of this during the experiment, though they were informed about the interaction logging beforehand. A first observation is that most of the participants did not follow the protocol in a strict linear fashion. Sometimes this was due to instructions, that have not been clearly understood, but became apparent on future steps. Though sometimes this was also because of slight delays in the voice recognition with Glass, leading to commands which were given twice. To still extract the currently active step, the interaction log was filtered to include only steps that were visible for more than a few seconds.

From this data, it is possible to extract which step of the DNA extraction experiment was viewed at which time, and the duration for which this instruction was visible on the display. Figure 5 visualizes the resulting workflow for each participant, as well as the mean interaction time per step. It can be seen that the overall interaction time ranges from 18 to 45 minutes for all participants, and the experiment was completed in 35 minutes on average. Furthermore, a larger break can be observed in the middle of the experimental workflow (cf. the two consecutive water bath steps), which matches the instruction from the experiment protocol: During these steps, participants had to wait until both the onion and tomato mixture had been cooled or heated up respectively, and no other experiment task could be performed during that period. While interpreting these figures it should be kept in mind that

the increasing variety in later steps is an artefact of the cumulative display of this particular step's duration. The figure to the bottom right contains the color-coded steps which are the same for both the tomato and onion extraction, i.e. the steps which are repeated for each participants. For example, preparing a solvent agent needs to be done twice, and is encoded in yellow in this figure. A large variety for solvent preparation time, in the duration of keeping the mixtures in the water baths, for filtrating and pestling can be observed.

Based on the data from these interaction logs in combination with video footage made during the evaluation study (for which a summary is depicted in Figure 5), the following three observations stand out especially:

- There is first of all a large cross-user variety concerning the duration of each step, most critically for the steps where participants were instructed to keep a fixed time (e.g. keeping the mixture in the water bath for a certain amount of time). For several participants, browsing through the steps using Glass' voice commands was too time-consuming and they switched to swiping gestures, mid-experiment. This large variety in performance times has as a consequence that timing within an experiment and duration of single steps are important parameters, though they are also less valuable for automatic detection or reconstruction of the experiment protocol.

- Even though participants were asked, for steps that consisted out of multiple items, to mark each item with a special voice command before the study, several found this too cumbersome and did not adhere to this instruction -

action	precision	recall	F1-score
cutting	0.23 ± 0.12	0.29 ± 0.17	0.25 ± 0.12
inverting	0.60 ± 0.21	0.64 ± 0.34	0.58 ± 0.25
peeling	0.04 ± 0.04	0.12 ± 0.15	0.06 ± 0.06
pestling	0.62 ± 0.21	0.42 ± 0.13	0.47 ± 0.12
pipetting	0.58 ± 0.13	0.52 ± 0.11	0.54 ± 0.11
pouring	0.03 ± 0.04	0.13 ± 0.27	0.04 ± 0.05
stirring	0.32 ± 0.25	0.37 ± 0.21	0.29 ± 0.15
transfer	0.08 ± 0.10	0.25 ± 0.34	0.08 ± 0.09
	0.31 ± 0.29	0.34 ± 0.29	0.29 ± 0.25

(a) *cross-participant*, i.e. leave-one-participant-out

action	precision	recall	F1-score
cutting	0.62 ± 0.20	0.75 ± 0.11	0.66 ± 0.16
inverting	0.80 ± 0.21	0.93 ± 0.16	0.84 ± 0.18
peeling	0.26 ± 0.17	0.51 ± 0.21	0.32 ± 0.17
pestling	0.87 ± 0.08	0.80 ± 0.07	0.83 ± 0.07
pipetting	0.79 ± 0.09	0.77 ± 0.06	0.78 ± 0.06
pouring	0.45 ± 0.22	0.80 ± 0.24	0.55 ± 0.21
stirring	0.83 ± 0.11	0.80 ± 0.08	0.81 ± 0.08
transfer	0.33 ± 0.28	0.49 ± 0.34	0.36 ± 0.28
	0.62 ± 0.29	0.73 ± 0.23	0.64 ± 0.26

(b) *per-participant*, i.e. detecting gestures for one participant only

Table 2. Precision/Recall/F1-Scores based on 250 iterations of a stratified random split (for per-participant scores) and leave-one-out cross-fold validation for cross-participant scores.

most often, these participants worked through the whole instruction set for one such a particular information card, and then marked all items in one go.

- Finally, it is important to note that all participants were able to finish their experiment with sole guidance of the wearable system, without abandoning the experiment, and extracting the DNA successfully. Figure 5 shows the time (in minutes) that all experiment steps took per participant.

Detecting Actions and Steps from Wrist-Movement

For evaluating the detection of experiment steps by means of the wrist-worn accelerometer data, the ground truth was gathered by annotating the recordings of an external camera[2], pointing at the manipulation area of the participant. By (manually) annotating the video we extracted 9 different actions, which had a high visual similarity and were repeated often during the experiment: The onion and the tomato were both *cut*, and the onion was also *peeled*. A pipette was used for combining different ingredients, e.g. *pipetting* the mixture into the test tube. The *transfer* activity describes using a spoon for putting, e.g. salt, in a beaker, but not using it for *stirring* for which a stirring rod was available. *Pouring* describes putting the mixture from one beaker to another or when pouring it into the filter. *Pestling* refers to mincing the mixture and *inverting* to putting the test tube upside down and back again. We refer to these video annotations as the ground truth in the following (cf. Table 2).

Wrist 3D-acceleration data was recorded throughout the experiment on the participant's dominant hand with a sampling rate of $50Hz$, within a range of $\pm 4g$. In total, $1258mins$ of accelerometer data were recorded. Additionally, the interaction with Google Glass was logged during the experiments, i.e. the timestamps when users switched to the next steps, and this data has been stored locally on the device and was aggregated on a PC after the experiment. For time synchronization we relied on the internal clocks of both Glass and the wrist-worn sensor. For video-annotated acceleration data, we additionally manually fine-tuned the alignment by matching the video and sensor data stream according to easily identified activities, such as stirring.

[2]The camera on Glass was not used to make sure that the whole manipulation area was visible throughout the experiment.

Action Detection

To recognize the activities listed in Table 2 from acceleration data, we chose a k-nearest Neighbour (k=8) classifier[3] with a 6D feature set, containing the mean and standard deviation of the 3D acceleration data during 20%-overlapping windows of $800ms$ duration. The video-annotated data was cross-validated *cross-participant* and *per-participant*. Cross-participant validation was achieved with a Leave-One-Participant-Out validation, while per-participant validation was done by a 250-times Stratified Shuffle Split. Precision, Recall, and F1-Scores for each evaluation are listed in Table 2. It is visible that cross-participant gesture recognition is much worse than per-participant: On average, the cross-participant F1-Score is 36% worse than per-participant, which is most probably caused by participants performing gestures in a slightly different fashion or due to sensors not being firmly attached. *Inverting*, *pestling* and *pipetting* are the three actions that show a particular high F1-score per-participant and comparable F1-scores across participants. In contrast to *stirring*, which is detectable per-participant but not cross-participant. *Cutting*, *peeling*, *pouring* and *transferring* (which in our case meant moving material with a spoon) are already hard to detect per-participant. The confusion matrices (cf. Figure 7) show that *pipetting* and *pestling* are most often confused with other actions, which therefore make it advisable to remove those if possible. From this we conclude that several characteristic actions can be detected with reasonable performance, when trained per-participant.

Protocol Step Detection

Although some actions can be detected well enough to allow a system to estimate a reconstruction of known experiment steps afterwards, it is unclear whether our proposed system would reliably detect single protocol steps. More specifically, it would be interesting to obtain accurate detections, from wrist movements alone, that show in which step in the experiment a participant is in. This can be an important cue when reviewing a video recording done in background and would allow the experimenter to jump to a specific point in the protocol during review. For investigating and evaluating this, we modeled the digitized protocol as a Hidden Markov Model (HMM), where hidden states correspond to steps in the experiment protocol, and observations map to the actions

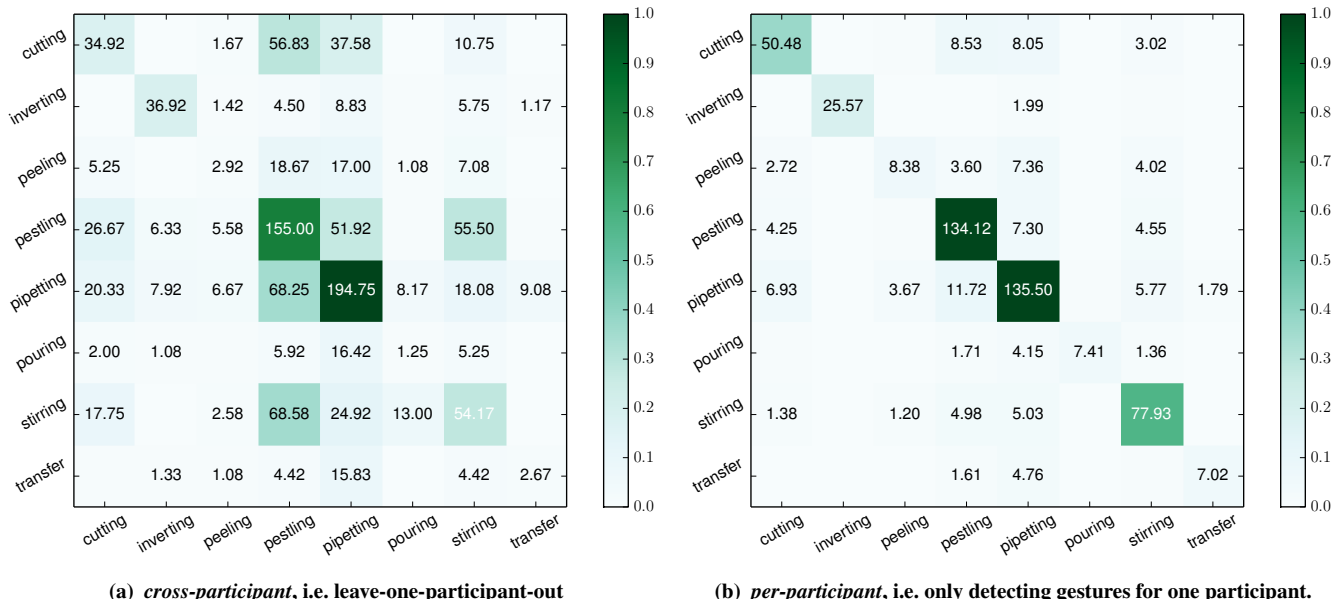[3]the scikit-learn [26] implementation was used.

Confusion matrix (a) — cross-participant, i.e. leave-one-participant-out

	cutting	inverting	peeling	pestling	pipetting	pouring	stirring	transfer
cutting	34.92		1.67	56.83	37.58		10.75	
inverting		36.92	1.42	4.50	8.83		5.75	1.17
peeling	5.25		2.92	18.67	17.00	1.08	7.08	
pestling	26.67	6.33	5.58	155.00	51.92		55.50	
pipetting	20.33	7.92	6.67	68.25	194.75	8.17	18.08	9.08
pouring	2.00	1.08		5.92	16.42	1.25	5.25	
stirring	17.75		2.58	68.58	24.92	13.00	54.17	
transfer		1.33	1.08	4.42	15.83		4.42	2.67

Confusion matrix (b) — per-participant, i.e. only detecting gestures for one participant.

	cutting	inverting	peeling	pestling	pipetting	pouring	stirring	transfer
cutting	50.48			8.53	8.05		3.02	
inverting		25.57			1.99			
peeling	2.72		8.38	3.60	7.36		4.02	
pestling	4.25			134.12	7.30		4.55	
pipetting	6.93		3.67	11.72	135.50		5.77	1.79
pouring				1.71	4.15	7.41	1.36	
stirring	1.38		1.20	4.98	5.03		77.93	
transfer				1.61	4.76			7.02

(a) *cross-participant*, i.e. leave-one-participant-out   **(b)** *per-participant*, i.e. only detecting gestures for one participant.

Figure 7. Confusion matrices for kNN detection based on $800ms$**-windowed mean and standard deviation features extracted from 3D-acceleration data. Cells contain the average absolute number of identified samples. The color designates the normalized total occurrence. Each stratified random split was repeated 250 times.**

executed in that steps as defined in Table 1. The Glass interaction log is split into equidistant timespans, which are marked with the current protocol step: This will serve as the ground truth and represents our detection target.

The HMM's transition probabilities of the hidden states were set to mimic a linear chain, with a high probability for staying in the same state (workflow step) and a non-zero probability to switch to the next step. This models the linear nature of a protocol execution. The emission probabilities for each state can be generated by uniformly distributing the occurrence of each action in the state's action set. For example, the *solvent* step (cf. Table 1) has high emission probabilities for the pouring, transfer, pipetting and stirring actions. The *detection* step in contrast only has a high probability for inverting. To account for possible mis-classifications of the kNN-detector, each action has a low occurrence probability in *each* workflow step. This represent a layered approach, in which the first layer detects actions from wrist movement via a kNN-detector and the second layer detects the workflow steps of the protocol via a HMM.
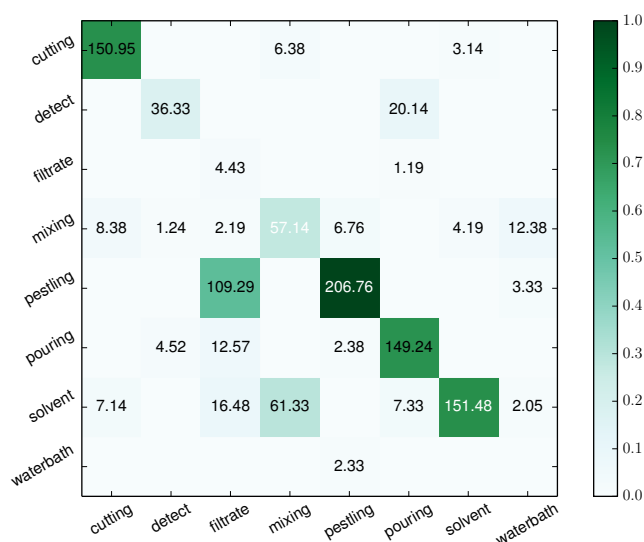
For the evaluation, we detected the workflow of each participant with the above-described kNN-HMM approach. We compared this workflow with the data gathered by the interaction log, i.e. which step was looked at when. Assuming that participants had the currently executed step also active on their display, we could also say that we check whether the *currently executed step* was detectable. The result of this evaluation can be seen in Figure 8. The confusion matrix shows that the *mixing* and *solvent* step are most often confused, which is due to the fact that they have almost the same action set. Only an additional *transferring* action distinguishes them, which is however hardly detectable by our kNN algorithm. With the presented layered approach, a mean F1-Score of 56% for detecting workflow steps from

wrist movements is achievable. This however includes workflow steps which have an empty action set, and are therefore difficult to detect. These steps include the *waterbath* and *filtrate* step, which for instance did not have definable activities linked to them: Excluding these steps from the calculation improves the mean F1-score to 71%. It is important to note, however, that such steps do occur in real wet lab experiments and therefore demand complementary detection approaches (e.g. object detection through RFID-markers) .

The presented layered approach can not only be used to filter wrist movement data on a time-based scale, but also to integrate different sources of information. Therefore, it lends itself well to integrating further sensors. One shortcoming of the presented HMM approach is that the actions set is assumed to be not ordered, i.e. it does not matter in which order the actions are executed. This might be important information, that is not directly modeled. In this case, conditional random fields (CRF) might prove to be a more suitable alternative. To be practical, a system like the one presented here would need to be either continuously re-learning motion sequences, or limit itself to actions that have proven to be detectable across users such as *inverting*, *pestling* and *pipetting*.

FUTURE WORK

As we now have a first implementation of a system that supports work flows in wet laboratory experiments, future work could integrate more sources of information and more means of support. One example is the use of object identification, for instance via RFID tags on the objects that are read out by a wrist-worn RFID reader [27] or by instrumented objects like a Bluetooth-enabled pipette. These features are also useful while establishing a protocol, e.g. instead of having to note down how much of a compound has been used, the pipette already "told" the system, similar to how the just used compound could be identified.

(a) mean *per-participant* confusion matrix

(b) mean *per-participant* classification scores.

#	task	precision	recall	F1-score
1	solvent	0.64 ± 0.17	0.93 ± 0.18	0.73 ± 0.15
2	cutting	0.88 ± 0.23	0.83 ± 0.28	0.84 ± 0.25
3	mixing	0.69 ± 0.31	0.49 ± 0.21	0.52 ± 0.18
4,5	waterbath	0.27 ± 0.37	0.09 ± 0.19	0.10 ± 0.20
6	pestling	0.64 ± 0.30	0.88 ± 0.25	0.72 ± 0.27
7	filtrate	0.62 ± 0.40	0.09 ± 0.21	0.12 ± 0.20
8	pouring	0.86 ± 0.23	0.79 ± 0.24	0.81 ± 0.22
9	detection	0.62 ± 0.36	0.75 ± 0.40	0.65 ± 0.35
		0.65 ± 0.35	0.61 ± 0.41	0.56 ± 0.36

Figure 8. Scores for the workflow step detection based on a Hidden-Markov Model, which uses observations from the kNN-based action classification. It is visible that classification scores vary between participants. Some steps are not detectable (waterbath, filtrating) since they have no definable and therefore detectable actions/observations (cf. Table 1).

The system could integrate many other features such as maintenance and resource allocation ("Is the centrifuge available for the next five hours?", "Is compound X in store?") across multiple lab members, and tools such as reminders for lengthier procedures or concentration calculators. Provisions to avoid cross-contamination through shared lab equipment (e.g. flasks, pipettes) could be extended by recording their usage, which would also allow tracing back contaminations after they have been detected. Retrieval and editing of the workflows after a completed experiment, would provide an extra possibility for the biologist to reflect on the results in detail. A recording infrastructure, such as the one presented in this paper, can be extended to provide a memory extension, detected actions, protocol steps, or used tools can serve as searchable cues for other recordings like video or audio. Furthermore, these cues might be employed to compare repetitions of the same protocol, allowing to quickly spot differences in their execution.

CONCLUSIONS

This paper presented a hands-free wearable system to support experimenters in a wet laboratory environment. A prototype, based on Google's Glass and a wrist-worn inertial data logger, was used to capture experiments and recall steps, navigate back and forth those steps, and mark them as done. An analysis of the challenges, as well as the acceptance and wearability of the system, based on "in-situ" observations in several different microbiology laboratories was conducted.

The original motivation of using Glass as a recording and guidance tool for an experimenter was tested for feasibility in the presented user study: 22 novice participants were asked to complete an interleaved entry-level DNA-extraction experiment. Participants solely relied on Glass for guidance on the procedure and were all able to finish their experiment successfully. Participants' wrist motions were

recorded throughout all experiments for studying whether actions made during experiments can be recognized, as well as used, for example in navigating continuous video recordings of procedures. Actions were detected with a k-Nearest-Neighbor classifier, of which only a limited set could be detected reliably (per-participant, F1-score > 80%). A Hidden Markov Model, built by extracting action sets for each protocol step from the digitized protocol, was used for detecting the currently executed step. This layered approach allowed to reconstruct the majority of experiment steps afterwards.

Detecting the executed steps in a protocol could serve as a cue for reviewing work post-experiment, or for automatically navigating a protocol on Glass while the experiment is performed. In this paper, we presented a benchmark dataset to evaluate detection strategies, with common actions found in many micro-biology experiments. Although our proposed system could be modified for use in other domains, e.g. following recipes in the kitchen, it is important to note that wet laboratories pose much stricter protocols. The ability to capture and review an experiment up to several weeks later is furthermore in wet labs more important than the actual guidance - when the experimeter finds out that something went wrong with the experiment, a detailed review of executed steps could shed light on the cause. Moreover, a wearable and touch-free system does not only decrease the chance of contamination, it also provides the means to interact with a computing system right on the spot, in turn minimizing the required interaction efforts.

REPRODUCIBILTY STATEMENT

To allow reproduction of the presented results, all video material and sensor recordings is publically available at **http://es.informatik.uni-freiburg.de/index.php/datasets/ubicomp2015/**.

REFERENCES

1. Eastmond, E. E., Mackay, W. E., Tabard, A., Eastmond, E. E., and Mackay, W. E. From Individual to Collaborative: The Evolution of Prism, a Hybrid Laboratory Notebook. *Comput. Support. Coop. Work.* 2008.

2. Bird, C. L., Willoughby, C., and Frey, J. G. Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences. *Chem. Soc. Rev.* 42, 2013: 8157–75.

3. Roubert, F., and Perry, M. Putting the Lab in the Lab Book : Supporting Coordination in Large , Multi-site Research. *Proc. 27th Int. BCS Hum. Comput. Interact. Conf.* 2002: 1–10.

4. Giles, J. Going paperless: The digital lab. *Nature*, 481(7382), Jan. 2012: 430–431.

5. Klokmose, C. N., and Zander, P.-O. Rethinking Laboratory Notebooks., 2010: 119–140.

6. Myers, J. D. Collaborative Electronic Notebooks as Electronic Records : Design Issues for the Secure Electronic Laboratory Notebook (ELN). *Simul. Ser.* 2003:

7. Taylor, K. T. The status of electronic laboratory notebooks for chemistry and biology. *Curr. Opin. Drug Discov. Devel.* 9(3), 2006: 348–353.

8. Bernstein, M., Kleek, M. A. X. V. A. N., Karger, D., and Schraefel, M. C. Information Scraps: How and Why Information Eludes our Personal Information Management Tools. *ACM Trans. Inf. Syst.* 26(4), 2008: 24.

9. Mackay, W. E., Pothier, G., and Letondal, C. The Missing Link: Augmenting Biology Laboratory Notebooks. *Symp. User Interface Softw. Technol.* 2002.

10. Yeh, R. B., Liao, C., Klemmer, S. R., Lee, B., Kakaradov, B., Stamberger, J., Paepcke, A., and Sciences, B. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. *Hum. Factors Comput. Syst.* 2006, 1–10.

11. Gemmell, J., Bell, G., Lueder, R., Drucker, S., and Wong, C. MyLifeBits: Fulfilling the Memex Vision. *Int. Conf. Multimed.* 2002.

12. Arnstein, L., Franza, R., Borriello, G., Consolvo, S., Hung, C.-y., Franza, R., Zhou, Q. H., Borriello, G., and Consolvo, S. Labscape: a smart environment for the cell biology laboratory. *IEEE Pervasive Comput.* 1(3), July 2002: 13–21.

13. Borriello, G. Invisible computing: automatically using the many bits of data we create. *Philos. Trans. A. Math. Phys. Eng. Sci.* 366(1881), 2008: 3669–3683.

14. Hughes, G., Mills, H., De Roure, D., Frey, J. G., Moreau, L., Schraefel, M. C., Smith, G., and Zaluska, E. The semantic smart laboratory: a system for supporting the chemical eScientist. *Org. Biomol. Chem.* Nov. 2004:

15. Coles, S. J., Frey, J. G., Bird, C. L., Whitby, R. J., and Day, A. E. First steps towards semantic descriptions of electronic laboratory notebook records. *J. Cheminform.* 5, 2013: 52.

16. Tabard, A., Hincapié-Ramos, J. D., Bardram, J. E., Ramos, J. H., and Bardram, J. E. The eLabBench in the wild: supporting exploration in a molecular biology lab. *Hum. Factors Comput. Syst.* 2012.

17. Echtler, F., Klinker, G., Häussler, M., and Klinker, G. BioTISCH: the interactive molecular biology lab bench. *CHI Ext. Abstr. Hum. Factors Comput. Syst.* 10, 5–10.

18. Scholl, P. M., and Van Laerhoven, K. Wearable digitization of life science experiments. *Proc. UbiComp '14 Adjunct*, 1381–1388.

19. Minnen, D., Starner, T., Ward, J. A., Lukowicz, P., and Tröster, G. Recognizing And Discovering Human Actions From On-Body Sensor Data. *Multimed. Expo, 2005. ICME 2005*, 2005:

20. Lukowicz, P., Timm-Giel, A., Lawo, M., and Herzog, O. WearITwork: Toward Real-World Industrial Wearable Computing. *IEEE Pervasive Comput.* 6, 2007: 8–13.

21. Nicolai, T., Sindt, T., and Witt, H. Wearable computing for aircraft maintenance: Simplifying the user interface. *Appl. Wearable Comput.* 2006.

22. Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., and Tr, G. Wearable Activity Tracking in Car Manufacturing. *IEEE Pervasive Comput.* 7(2), Apr. 2008: 42–50.

23. Ockerman, J., and Pritchett, a.R. Preliminary investigation of wearable computers for task guidance in aircraft inspection. *Int. Symp. Wearable Comput.* 1998: 33–40.

24. Berlin, E., and Van Laerhoven, K. Detecting leisure activities with dense motif discovery. *Proc. UbiComp '12*, 250–259.

25. Gruber, J. Markdown: Syntax. address: `http://daringfireball.net/projects/markdown/syntax`.

26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2011:

27. Berlin, E., Liu, J., Laerhoven, K. V., and Schiele, B. Coming to Grips with the Objects We Grasp: Detecting Interactions with Efficient Wrist-Worn Sensors. *Tangible, Embed. Embodied Interact.* 2010.