1. Introduction

Pair programming (PP) is a software development practice in which two developers work closely together on a technical task on the same computer. It was popularized by Kent Beck who sees it as the central practice of eXtreme Programming and describes it as “a dialog between two people trying to simultaneously program (and analyze and design and test) and understand together how to program better” [2, p. 100].

Controlled experiments on pair programming have shown mere tendencies in terms of effects on quality and effort with much variation left to be explained [3]. In the words of the authors of a large experiment with almost 300 hired consultants: “we are still far from being able to explain why we observe the given effects” [1].

Our research group has been collecting industrial pair programming sessions since 2007. We record pair programming as it happens “in the wild” in order to understand how it actually works and what really matters in everyday practice. In particular, we record the pairs’ conversation, their screen content, and a webcam video showing their gestures and posture.

This kind of data is amendable to different types of analyses. We describe our qualitative approach in [10, 11]. In this report, we describe the technicalities of how we collected the data. Several researchers have contributed plenty of time to collecting and processing that data, and we want to give credit where credit is due. The raw data itself cannot be released to the public because of non-disclosure agreements with the respective companies. As a proxy, we here characterize the companies, the developers, and their PP sessions.

This report is structured as follows: We discuss our fundamental approach to collection empirical data on pair programming (Section 2) and describe our generic data collection protocol (Section 3). We introduce some terminology and describe the structure of our data (Section 4). We give an overview of our repository (Section 5) and then discuss the individual contexts and cases in (Section 6). We close with a discussion of the properties and limitations of our data collection (Section 7) and give an overview of which data has been used in which publications so far (Section 8). In Appendix A, we explain the technical details of how we record and process PP sessions; Appendix B is a (redacted) reprint of a report we handed out to one of the companies.
2. Fundamental Considerations

There are two fundamental considerations to our data collection: First, we record pair programming as it happens in industry. Second, we consider the pair programming session as the basic unit. Both considerations were driven by our research interests.

2.1. Naturalistic Industrial Setting

Our research has a focus on practical relevance. This why we study industrial settings with professional software developers working on their everyday tasks. This also entails that the developers work in their normal development environment, with partners they chose to work with, at times and to an extent they decide.

We primarily rely on observation of developers working in pairs, as opposed to reports of pair programming in, say, interviews. To enable a thorough analysis later on, we record the pair programmers. In particular, the pair members’ interactions with one another and their computer(s) as well as the contents of their screen(s) need to be captured in audio and video. The necessary recording infrastructure somewhat reduces the naturalism of the observed session, the effects of which we discuss in Section 7.2.

2.2. The Pair Programming Session as a Unit

Our data collection starts a point where the developers already made the decision to work as a pair. Their decision, just as the project they work in, the task(s) they work on, their software system, and their team structure all may “echo” in their session and are therefore helpful for understanding their activities, but are not prime concerns of our data collection.

This constrasts, for example, with the work of Socha et al. [15, 16] who recorded 11 days of software development with a multitude of video cameras and audio recorders resulting in six terabytes of data consisting of thousands of photographs, 292 hours of screen captures, and 380 hours of video of developer interactions.

3. Data Collection Protocol

In our research group, Stephan Salinger and Laura Plonka initiated our industrial data collection efforts and they devised a protocol that served as the basis for collecting data in all companies.

The data collection protocol is generic in two ways. First, it is adapted in each particular installment at a company on-site to deal with constraints, to seize opportunities, and to fit the particular research focus of the researcher (see Section 3.3). Second, the protocol is still more or less independent from any particular research question regarding pair programming, as the resulting data can be reused for different purposes (some conditions apply, which we discuss in Section 7).

3.1. Protocol Overview

After a company has been approached and probed whether the company would be open to have some of their programming sessions recorded, the overall research goal, the procedure, extent, and purpose of the main data collection are explained in a presentation for the development team(s). It is also made clear to the potential subjects that all participation is voluntary and that their agreement to be recorded can be revoked at any point during a session. The heart of the data collection protocol is a cycle per session recording:

- After a pair announces that it is willing to have their next pair session recorded, the recording infrastructure is set up and the session recording is started once the developers are ready (see Section 3.2 for details).
- Optionally, both developers fill out questionnaires before and/or after their session, in which the developers state their names, development and pair programming experience, characterize the nature of their task, and whether it went as they intended (see Fig. 2).
- Afterwards, the researcher does a quick analysis of the material during which she looks for peculiarities that catch her attention. The main purpose of this step is to inform the next activity.
- The researcher then conducts a reflective interview with the developers on the day after the recording. This activity serves different purposes, including collecting background information and providing developers with feedback in return for them agreeing to have their PP session recorded and scrutinized. These interviews are audio-recorded.

3.2. Recording Sessions

The software developers themselves decide when and for how long they want to have their work recorded. They work on their own machines, in their normal environment, on their everyday tasks, and with the partner they chose.

The session recordings as technical artifacts consist of a screencast of the pair’s monitor(s), the pair’s conversation as audio, and a webcam video showing the two pair members’ interaction. These three sources are combined to a self-contained video file as illustrated in Fig. 1. Both webcam feed and screencast are captured at between 5 and 15 frames per second (depending on what the hardware can handle), which does not make for a great movie experience but allows to distinguish individual keystrokes, to follow mouse movements, and to see the developers’ gestures. The final videos’ resolution depends on the developers’ display(s) and recording setup and ranges from 1024×768 to 2560×1440 pixels.

The recording process relies on one of three generations of hardware and software components. The general setup works like this: The developers work on one machine, and screencast and webcam feed are transmitted to another machine where they are recorded. We explain the details in Appendix A. The most relevant difference is that generation 1 is an unattended offline recording which the researcher only gets to see once the pair is done, while generations 2 and 3 are an online recording which allows the researcher to watch the session live (and start her quick analysis) while it is still being recorded.
3.3. Per-Company Differences

With each installment of the data collection protocol at a new company, there were slightly different sets of mutual expectations which resulted from prior discussions with the partners and from evolved research interests. They can be grouped into three areas as to what the “headline” of the cooperation was. We discuss this difference because it affects the researcher’s emphasis she puts on each of the data collection activities on-site. Additionally, the subjects’ expectations and possibly also behavior are shaped by knowing why a researcher is around. Table 1 gives an overview of the contexts, involved researchers, and headlines.

- **PP**: Most companies were specifically approached with the intention to understand pair programming. While for the first contacts there was no particular focus yet, Plonka’s [4] contacts to companies C to F were influenced by a particular interest in the *driver* and *navigator* roles, which is why the webcams in these sessions are recorded in an angle that shows the developers hands but occasionally cuts off their heads. Additionally, the data collection period was branded to these companies as a “workshop” to help developers reflect on their PP process. Here, one workstation was setup for all pairs to use and the developers would put their names on a list to choose either a morning or an afternoon slot to be recorded for a planned maximum duration of 1.5 to 2 hours. Nevertheless, once a recording was started, the developers (just as in all other companies) were left to work on tasks of their choosing for as long they wanted. Schenk [13] was particularly interested in distributed pair programming and chose her contacts J and L accordingly; Zieris [17, 18] is interested in knowledge transfer and all participants from companies K, M, O, and P knew that.

- **Agile**: Salinger and Zieris approached companies K and O in an effort to understand agile software development in general. According data from these contexts was collected and analyzed (see [20], which is about company K). All particular PP sessions, however, were recorded exclusively for the purpose of understanding knowledge transfer in pair programming.

- **Onboarding**: Schmeisky, Salinger, and Zieris approached company N for understanding their onboarding process, i.e., how new-hires are integrated into the company. Pair programming was not a designated part of that process, but a number of developers agreed to be recorded while working in pairs. The recordings were therefore a window into the actual onboarding
process, and were not made to understand pair-programming.

Another effect of these different headlines and focuses can be seen in the different versions of the pre- and post-session questionnaires, see Figure 2.

4. Terminology and Structure

We use the following nomenclature:

- Companies are represented by single letters: A, B, C, and so on.
- Sessions are grouped by their technical context (to distinguish different sets of requirements and/or differences in technology stacks) and then counted up with Arabic numbers. CB1 is the first recording in the second context in the third company.
- Developers are identified through their company and Arabic numbers, such as C3.

4.1. Pair Programming Modes

Although our primary research focus is on pair programming, we did not restrict our recordings to a fixed settings of two developers working on the same machine. We distinguish three dimensions along which the collaboration modes in our repository differ.

- **Number of Developers**: Just as one developer may ask a colleague for help, a pair may do the same. Some teams even form groups of three or more people intentionally. So far, we recorded sessions with groups of two to four developers.
- **Spatial Distribution**: The developers can all be co-located, or all be in separate locations. (There may also be co-located subgroups, but we did not record such settings yet.)
- **Number of Active Computers**: The developers can share access to a single machine, or each of them can use their own machine, or anything in between.

In principle, these three dimensions are independent from each other. However, we did not observe every possible constellation yet. We characterize the ones we did see as follows (refer to Table 2 for a summary):

- **Pair Programming (PP)**: Two developers sit next to each other and work on one single machine.
- **Pair Programming with Active Observer (PPao)**: Like PP, but one partner occasionally looks up things on her own machine. The pair still works on one task though.
- **Side-by-Side Programming (SbS)**: Two developers sit next to each other and work on their own separated but related tasks.
- **Mob Programming (Mob)**: Three or more developers sit close to each other (e.g. in a conference room next to each other in a row) and work on anything between one single machine, and one machine per developer. We have observed neither fully nor partially distributed Mob settings.
- **Remote Pair Programming (RPP)**: Two developers are distributed to two different locations and share one screen, i.e., there is one developer who “owns” the machine, and the other developer may have view-only access (screen-sharing, e.g., via Skype) or some interaction options (e.g., via TeamViewer).
- **Distributed Pair Programming (DPP)**: Two developers are distributed over two different locations and they each can interact with their development environment independently but their actions are synchronized.

4.2. Structured Developer Information

For all 57 developers, we provide the following structured information:

- **Gender (Gnd)**: We did not ask any participant for their gender specifically. We list them as either “female” or “male” depending on the first name they told us.
- **Spoken Languages (Lang)**: We list the developers’ spoken languages in decreasing proficiency, using ISO 639-1 codes (e.g., “EN” for English, “DE” for German). We only list languages that are actually spoken in our data; a dash “–” is used as a placeholder, e.g., when it is clear that the session language is not the developer’s first language and she never uses that first language.
- **Software Development Experience (Dev.)**: In some companies, we asked the developers to self-report their professional software development experience in years and months in the questionnaires (see Fig. 2). In other cases, we retrospectively searched for public profiles on professional network sites.
- **Pair Programming Experience (PP)**: Here, we only relied on self-reported data from the questionnaires.
- **Time with Company (Comp.)**: This aspect we did not ask for in any questionnaire. The according numbers stem exclusively from the participants’ public profiles on career sites.

For all the experience-related information we use the following conventions: A “n.a.” in a table cell means that no information whatsoever is available. A blank cell means that the developer did fill out the according questionnaire, but left the field blank. Normal numbers are represented as such, while special markings and comments are given in quotation marks, e.g. “ca. 6?” or “first time”. If the same developer was asked the same question on multiple occasions and gave different answers, the respective values are separated by a slash “/”.

- **Number of Sessions (#S)**: The number of session recordings in our repository the developer is part of.
- **Number of Pairs (#P)**: The number of distinct pair (or group) constellations in which the developer was recorded.

4.3. Structured Session Information

For all 67 sessions, we provide the following structured information:

- **Mode**: A characterization of how the developers work together (see Section 4.1). We list the predominant mode of the session. See the tables’ respective captions and Section 6 for more details on some corner cases.
Table 1: Contexts for sessions recorded by member of our research group. The research direction (“headline” and “focus”) was set by the named researchers. Sessions in three additional industrial contexts (G, H, and I) were recorded with low technical quality and by students with little oversight, so important context information is missing. We do not include the according sessions in our repository PP-ind because of their low quality.

<table>
<thead>
<tr>
<th>Company</th>
<th>Application Domain</th>
<th>Country</th>
<th>Year</th>
<th>Researchers</th>
<th>Headline</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Content Management System</td>
<td>Germany</td>
<td>2007</td>
<td>Salinger</td>
<td>PP</td>
<td>–</td>
</tr>
<tr>
<td>B</td>
<td>Social Media</td>
<td>Germany</td>
<td>2014</td>
<td>Schenk</td>
<td>PP distributed PP</td>
<td>–</td>
</tr>
<tr>
<td>C</td>
<td>Geo-Information System</td>
<td>Germany</td>
<td>2013</td>
<td>Plonka</td>
<td>PP workshop, roles</td>
<td>–</td>
</tr>
<tr>
<td>D</td>
<td>Customer Relationship Management</td>
<td>Germany</td>
<td>2013</td>
<td>Salinger &amp; Zieris</td>
<td>Agile &amp; PP</td>
<td>knowledge transfer</td>
</tr>
<tr>
<td>E</td>
<td>Logistics and Routing</td>
<td>Germany</td>
<td>2007</td>
<td>Salinger</td>
<td>PP</td>
<td>–</td>
</tr>
<tr>
<td>F</td>
<td>Email Marketing</td>
<td>Germany</td>
<td>2016</td>
<td>Zieris</td>
<td>Agile &amp; PP</td>
<td>knowledge transfer</td>
</tr>
<tr>
<td>J</td>
<td>Data Management for Public Radio Broadcast</td>
<td>Germany</td>
<td>2013</td>
<td>Schenk</td>
<td>PP distributed PP</td>
<td>–</td>
</tr>
<tr>
<td>K</td>
<td>Estate Online Platform</td>
<td>Germany</td>
<td>2018</td>
<td>Zieris</td>
<td>PP</td>
<td>–</td>
</tr>
<tr>
<td>L</td>
<td>Freelance Training Consultant</td>
<td>USA</td>
<td>2016</td>
<td>Salinger, Schmeisky, Zieris</td>
<td>Onboarding</td>
<td>–</td>
</tr>
<tr>
<td>M</td>
<td>Data Analysis in Energy and Transportation</td>
<td>Norway</td>
<td>2014</td>
<td>Schenk</td>
<td>PP distributed PP</td>
<td>–</td>
</tr>
<tr>
<td>N</td>
<td>Online Fashion Retailer</td>
<td>Germany</td>
<td>2016</td>
<td>Zieris</td>
<td>Agile &amp; PP</td>
<td>knowledge transfer</td>
</tr>
<tr>
<td>O</td>
<td>Online Project Planning</td>
<td>Germany</td>
<td>2018</td>
<td>Zieris</td>
<td>PP</td>
<td>–</td>
</tr>
<tr>
<td>P</td>
<td>Online Car Part Resale</td>
<td>Germany</td>
<td>2016</td>
<td>Zieris</td>
<td>Agile &amp; PP</td>
<td>knowledge transfer</td>
</tr>
</tbody>
</table>

Table 2: Collaboration modes in our data collection

<table>
<thead>
<tr>
<th>Name</th>
<th>Short</th>
<th>#Dev</th>
<th>Spatial Setting</th>
<th>Active Computers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair Programming</td>
<td>PP</td>
<td>2</td>
<td>co-located</td>
<td>1</td>
</tr>
<tr>
<td>Pair Programming with Active Observer</td>
<td>PPao</td>
<td>2</td>
<td>co-located</td>
<td>1–2</td>
</tr>
<tr>
<td>Side-by-Side Programming</td>
<td>SbS</td>
<td>2</td>
<td>co-located</td>
<td>2</td>
</tr>
<tr>
<td>Mob Programming</td>
<td>Mob</td>
<td>3+</td>
<td>co-located</td>
<td>1</td>
</tr>
<tr>
<td>Remote Pair Programming</td>
<td>RPP</td>
<td>2</td>
<td>distributed</td>
<td>1</td>
</tr>
<tr>
<td>Distributed Pair Programming</td>
<td>DPP</td>
<td>2</td>
<td>distributed</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Collaboration modes in our data collection

- **Developers**: An incident matrix of the developers participating in the session.
- **Start**: Date and time when the session proper started. For some recordings, the video starts several minutes earlier.
- **Duration (Dur.)**: Gross length from the start to the point when the pair/group dissolves their session, including any short breaks. An intermittent stand-up meeting of 15 minutes or less counts as short whereas a lunch break does not in which case we consider the parts to be individual sessions even if they where recorded in one sitting.

- **Spoken language(s) (SL)**: The natural languages used by the participants in the session, using ISO 639-1 codes (e.g., “EN” for English, “DE” for German).

- **Programming Language(s) (PL)**: The structured languages of the source code files/snippets the developers received and answered the task & pair items (except 4, 6, and 9) via e-mail after their respective sessions. In the other installments (M, N, O, P), no questionnaires were used. Instead, the researchers asked the developers directly in the reflective interviews.

- **Pre-Session Pair Assessment (Pre)**: Answer to task & pair item 8: ‘How well attuned to your partner are you?’ expressed in German school grades from 1 (best) to 6 (worst).

- **Post-Session Assessment (Post)**: Answer to process item 1: ‘How useful was it to solve the task as a pair?’ expressed in German school grades from 1 (best) to 6 (worst).
5. Overview of Sessions

Table 3 summarizes the contents of the repository per company. In addition to the types and lengths of the recorded sessions, we also list the quality of the auxiliary data sources, i.e., pre- and post-session questionnaires and recordings of the reflective interviews.

Figure 3 is an histogram of the sessions’ respective lengths. The vast majority of our sessions is runs between 0:45 and 2:30 hours.

Attached to this arXiv publication is an R file that contains all the structured information in the form described in Sections 4.2 and 4.3.

6. The Repository

In this Section, we characterize each of the 13 companies and provide summaries for many of the individual sessions. Analyzing such sessions on a content-level is hard work and takes time. Hence, we only provide summaries for the sessions whose technical contents we already understand good enough.

6.1. Company A

Company A develops a web-based content management system (CMS) in Java and Objective-C.

We recorded a single session (AA1) with developers A1 and A2 in this company. Both developers know their domain well and are generally experienced developers. Developer A1 has very good structural knowledge of the Java frontend and its individual classes, as well as practical knowledge of the Eclipse IDE and the Java programming language. His colleague A2 is more familiar with the backend and the SQL database, the VIM editor and the UNIX shell, and the Objective-C programming language. However, these differences are small: Each of them would be able to find their way around in both parts of the system.

See Table 4 for general information on the developers and Table 17 for structured information on the recorded session.

Session AA1 The pair works on fixing inconsistencies across different list views. They work both in the frontend and the backend code. Although they know their code base well, they still spend some time along to way to understand its peculiarities. At the end of the session they worked through five different lists and made them consistent, but learned that one aspect of the business model cannot be implemented in the backend alone. They finish with a architectural discussion.

6.2. Company B

Company B develops a social media platform in PHP and JavaScript. We visited the company two times to record a total of four PP sessions.

All four sessions are with the same pair of full-stack developers B1 and B2. See also Salinger’s description of the data collection [9, pp. 95–99]. See Table 5 for general information on the developers and Table 18 for structured information on the recorded sessions.

Note on Naming Scheme: We originally considered the data from the first recording sitting to be corrupt but were eventually able to mostly recover them. This is why the earlier sessions are labeled BB1 to BB3 even though they take place five months prior to session BA1.

Sessions BB1, BB2, and BB3 The pair implements a new feature from scratch in the course of one afternoon (half past two until half past eight), going through their complete web development stack, starting with template and internationalization in session BB1, continue with controller, model, database layer, and template optics in session BB2, and conclude with making their view more interactive through JavaScript in session BB3.

Note on Data: The webcam and audio recording is jumbled for the last 30 minutes of session BB3: While the screen video is continuous, about 10 minutes worth of webcam and audio are randomly missing, making the other 20 minutes very difficult to understand due to lack of continuity.

Session BA1 The pair takes over some code of unknown quality written by outsourced developers. Technically, they want to implement part of a cache. In particular, their logic should tell whether the requested data has changed since a given timestamp. In their session, they need to understand all existing code for that functionality (a few dozen lines of PHP code), make some additions, and encounter difficulties in specifying what exactly their cache should do. In the first minutes they also struggle with the workstation which is not theirs and not fully configured to their needs.

6.3. Company C

Company C develops a geographic information system desktop GUI application written in Java. The design of this software uses abstraction elaborately.

Over the course of one week, we recorded 6 sessions involving 8 developers in this company. See also Plonka’s description of the data collection [4, pp. 64–67], and the (German) handout produced for company C in Appendix B. See Table 6 for general information on the developers and Table 19 for structured information on the recorded sessions.
Developers C1 and C2 work on a new form on the system’s GUI. C1 has started working on the form prior to the session; C2 is new to the task. In their session, they mostly deal with making their new GUI component in the form toggleable for which they reuse their session, they mostly deal with making their new GUI form prior to the session; C2 is new to the task.

Session CA2 Developers C2 and C5 work on a small functional extension and its main difficulty lies in understanding and properly applying the existing design abstractions. The task was started by C5 prior to the session. The work consists of some design-discussion and abstractions. The task was started by C5 prior to the session; C2 is new to the task. In their session, they mostly deal with making their new GUI component in the form toggleable for which they reuse existing GUI logic.

Session CA3 Developers C6 and C7 want to implement a new context menu entry which is only enabled under certain circumstances. They write test cases for the menu entry to be enabled and disabled, and refactor code along the way. This is a simple task, but their IDE freezes many times for over a minute each time which slows the pair down a lot. After 1:20 hours into their session, the pair took a break of four minutes.

Session CA4 Developers C4 and C7 implement a new feature that allows users to cut existing geometries into parts by drawing a new feature that allows users to cut existing geometries into parts by drawing arbitrary shapes across them. The pair also uses two sets of keyboard and mouse quite some time is lost in the second half of the session due to a problem with the team’s SVN (Subversion) server. The pair also uses two sets of keyboard and mouse fluently.

Session CA5 Developers C3 and C4 start implementing a new feature that allows users to cut existing geometries (such as points, lines, polygons) into parts by drawing arbitrary shapes across them. The pair uses two sets of keyboard and mouse fluently and generally have a productive and high-paced session.

Session CB1 Developers C4 and C8 work on a pet project of theirs before official office hours. It is written in Java, too, but has nothing to do with the domain of the C company.
6.4. Company D

Company D develops a large customer-relationship system that is based on Eclipse and comprises some 50 top-level modules written in Java.

During one week, we recorded 6 sessions involving 8 developers in this company. See also Plonka’s description of the data collection [4, pp. 64–67]. See Table 7 for general information on the developers and Table 20 for structured information on the recorded sessions.

![Table 7: Overview of the D developers. See Section 4.2 for information on the data and its representation.](image)

### Session DA1

**Note on Data:** No webcam was recorded.

### Session DA2

Developer D4 is in his very first week at the company and has never pair programmed before; his colleague D3 has been with the company for three months. It is his first programming job for which he started learning Java.

In their session, they try to implement a new toolbar for one of the system’s many modules. After some failed attempts and two long discussions with additional developers (first D7, then D6), they perform a technically simple refactoring task which spans many different modules and lasts the rest of session. Throughout the session, D4 provides D3 with information on programming styles, technologies, and so on, whereas D3 is more knowledgeable about the code base and the organizational background.

The team had its daily stand-up meeting in the middle of the session which leads to a 15 minute break.

6.5. Company E

Company E develops a graphical desktop application for different logistics related tasks in the C++ programming language, but there are also parts written in C# and Java.

During one week, we recorded 7 sessions involving 8 developers in this company. See also Plonka’s description of the data collection [4, pp. 64–67]. See Table 8 for general information on the developers and Table 21 for structured information on the recorded sessions.

### Session EA1

Prior to the session, developer E2 already tried to debug a display error that leads to routes of ferries being displayed with an extra segment. In their PP session, he and colleague E1 go through the unfamiliar source code step by step with a debugger. They repeatedly set up a certain state, inspect variables, develop and test hypotheses, but they do not change any code. They end their session after 80 minutes because of a team meeting without having really circled in on the error.

**Note on Data:** There are several pauses in the video stream because the developers accidentally paused and resumed the recording underway by using the F9 and F10 keys in their debugger which the recording software also listened to.

6.6. Company F

Company F develops an email marketing tool written in Java.

On three consecutive days, we recorded 4 sessions involving 6 developers in this company. See also Plonka’s description of the data collection [4, pp. 64–67]. See Table 9 for general information on the developers and Table 22 for structured information on the recorded sessions.

### Session FA1

Prior to the session, developer F1 already tried to debug a display error that leads to routes of ferries being displayed with an extra segment. In their PP session, he and colleague F2 go through the unfamiliar source code step by step with a debugger. They repeatedly set up a certain state, inspect variables, develop and test hypotheses, but they do not change any code. They end their session after 80 minutes because of a team meeting without having really circled in on the error.

**Note on Data:** There are several pauses in the video stream because the developers accidentally paused and resumed the recording underway by using the F9 and F10 keys in their debugger which the recording software also listened to.

### Session FA2

Developer F3 is in his very first week at the company and has never pair programmed before; his colleague F4 has been with the company for three months. It is his first programming job for which he started learning Java.

In their session, they try to implement a new toolbar for one of the system’s many modules. After some failed attempts and two long discussions with additional developers (first F7, then F6), they perform a technically simple refactoring task which spans many different modules and lasts the rest of session. Throughout the session, F4 provides F3 with information on programming styles, technologies, and so on, whereas F3 is more knowledgeable about the code base and the organizational background.

The team had its daily stand-up meeting in the middle of the session which leads to a 15 minute break.

6.7. Context J

Company J is a service provider for public radio broadcasters. All recorded PP sessions involve the same pair of developers who do **distributed pair programming** because they work in different cities: J2 is directly employed by said company and J1 is a consultant. The pair synchronized their development environments with the Eclipse plugin “Saros”, which allows concurrent editing in all shared files for both developers.

Their 9 sessions were recorded on four days: Session JA1 on one day, sessions JA2 to JA9 two weeks later on

---

1. Project homepage: https://www.saros-project.org
three consecutive days. See also Schenk’s description of the data collection [13, pp. 137–139].

J1 and J2 have known each other for about one year, the first two months of which they shared an office. However, as a consultant, J1 first got in contact with the concrete application domain 8 months prior to the session recordings. See Table 10 for general information on the developers and Table 23 for structured information on the recorded sessions.

Table 10: Overview of the J developers. See Section 4.2 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gnd</th>
<th>Lang</th>
<th>Dev. Y M</th>
<th>PP Y M</th>
<th>Comp. Y M</th>
<th>#S #P</th>
</tr>
</thead>
<tbody>
<tr>
<td>J1</td>
<td>male</td>
<td>DE</td>
<td>6 4 n.a.</td>
<td>0 8</td>
<td>1 6</td>
<td>9 1</td>
</tr>
<tr>
<td>J2</td>
<td>male</td>
<td>DE</td>
<td>1 6 n.a.</td>
<td>1 6</td>
<td>1 6</td>
<td>9 1</td>
</tr>
</tbody>
</table>

Session JA1 Domain expert J2 had designed and implemented a plugin-based architecture in Java to monitor and download remote files from the servers of different radio stations about a year earlier. He invites experienced consultant J1 to a distributed pair programming session to review and clean-up the code together in order to ease the implementation of a new feature later on. In the session, they review but one class, try and fail to refactor it by extracting local methods, and ultimately decide to rewrite the whole system from scratch, which they do two weeks later in sessions JA2 to JA9.

Session JA2 The pair starts with developing the module from scratch. In the first part of the session, J2 shows J1 a number of helper implementations he wrote in the meantime and J1 criticizes them. Afterwards, they discuss and collect requirements together in a plain text file.

6.8. Company K

Company K develops and operates a large web-portal for many estate-related services using different technologies which are separated through a microservice architecture.

We collected data on multiple occasions: Sessions KA1 and KA2 show an inter-team collaboration, sessions KB1 and KB2 show in-team work two months later, and KC1 to KC4 take place yet another six months later after the team had changed its technology stack. See Table 11 for general information on the developers and Table 24 for structured information on the recorded sessions.

Note on Naming Scheme: Originally, we considered all 8 K-sessions to be of the same context and numbered them accordingly “KA1” to “KA8”. However, we later decided to split them into three different contexts due to shift in involved systems and even programming languages. Therefore, when we mention “session KA6” in [18], we actually refer to session KC2.

Session KA1 and KA2 Developers K1 and K2 come together to work out an API between their respective teams’ subsystems: K1 is responsible for a mobile app for which K2 writes the endpoint with Java Spring web framework.2 Before they can start with their actual task, they first need to change the target URL of a single link which takes them more than 45 minutes and the help of two colleagues, because their development environment was not properly set up. Afterwards, K1 explains the data he needs with some dummy JSON file he prepared and K2 considers which internal microservices are able to provide which kind of data.

After a lunch break, the continue with a first implementation in session KA2. Overall, their pair work involves reading a lot of somewhat-known source code and existing API specifications, and generates a lot of fresh common ground between the two.

Sessions KB1 and KB2 These two sessions were recorded two months after session KA1 and KA2. Developers K2 (who is now more experienced in the domain) and K3 (who is more of a database expert) amend their data model: First they introduce a new model class and discuss which fields to include. In the second half they write and debug a database migration to adapt the database schema.

Sessions KC1 and KC2 Yet another six months later, developers K2 and K3 work together in their now changed environment: The team switched its technology stack from Java to CoffeeScript.3 The two are in the process of getting to know the jQuery JavaScript library4 because they want to write an integration test of an auto-completion feature, for which they want to programmatically enter characters into an input field. In session KC1, they set up their test environment and discuss different test approaches. In session KC2, after a lunch break, they try out these approaches which does not work out as intended and they struggle with their debugger.

6.9. Context L

L1 is a freelance consultant who offers training sessions to individual developers who bring their own tasks to work on with him remotely.

We recorded two of these training sessions with two different clients L2 and L3. See Table 12 for general information on the developers and Table 25 for structured information on the recorded sessions.

Session LA1 Consultant L1 and developer L2 go through a tutorial on the Ruby language.5 L1 is proficient in Ruby (but still learns some fundamental aspects) whereas L2 is fairly new to it.

Table 11: Overview of the K developers. See Section 4.2 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gnd</th>
<th>Lang</th>
<th>Dev. Y M</th>
<th>PP Y M</th>
<th>Comp. Y M</th>
<th>#S #P</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>male</td>
<td>DE</td>
<td>0 6 n.a.</td>
<td>0 6</td>
<td>2 1</td>
<td></td>
</tr>
<tr>
<td>K2</td>
<td>male</td>
<td>DE</td>
<td>n.a. 2 6</td>
<td>0 3</td>
<td>8 3</td>
<td></td>
</tr>
<tr>
<td>K3</td>
<td>male</td>
<td>DE</td>
<td>8 2 / 3</td>
<td>1 1</td>
<td>4 1</td>
<td></td>
</tr>
<tr>
<td>K4</td>
<td>male</td>
<td>DE</td>
<td>12 n.a.</td>
<td>1 3</td>
<td>2 1</td>
<td></td>
</tr>
</tbody>
</table>

2. Project homepage: https://spring.io/
3. Project homepage: https://coffeescript.org/
4. Project homepage: https://jquery.com/
5. Tutorial homepage: http://rubykoans.com/
Developers N1 and N2 sit next to each other and work through the documentation, but each of them configure their own machine. They take a break for about one hour between NA2 and NA4.

**Note on Data:** While N3’s audio, screen, and webcam are recorded, for her partner N4 only the audio is available.

**Session NA5** One week later, developers N2, N3, and N4 sit together at one table, each of them still trying to set up their respective development environments. They exchange ideas and insights for three and a half hours.

**Note on Data:** N4’s webcam is not recorded, but all other channels are (i.e., 3× audio, 3× screencast, and 2× webcam).

### 6.12. Company O

Company O develops a web-based project planning tool using CoffeeScript in both frontend and backend.

During a four-week observation period we recorded 10 sessions on 4 recording days. The team employed all of normal pair programming, mob programming, side-by-side programming, and remote pair programming. They also spoke both German and English. See Table 15 for general information on the developers and Table 28 for structured information on the recorded sessions.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gnd</th>
<th>Lang</th>
<th>Dev. Y M</th>
<th>PP Y M</th>
<th>Comp. Y M</th>
<th>#S #P</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>male</td>
<td>DE, EN</td>
<td>10 8</td>
<td>n.a.</td>
<td>1</td>
<td>7 4</td>
</tr>
<tr>
<td>O2</td>
<td>male</td>
<td>DE, EN</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0</td>
<td>5 2 1</td>
</tr>
<tr>
<td>O3</td>
<td>female</td>
<td>-, EN, DE</td>
<td>1 1</td>
<td>n.a.</td>
<td>0</td>
<td>1 9 5</td>
</tr>
<tr>
<td>O4</td>
<td>male</td>
<td>DE, EN</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0</td>
<td>5 9 5</td>
</tr>
<tr>
<td>O6</td>
<td>male</td>
<td>DE, EN</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>1 1</td>
</tr>
</tbody>
</table>

**Pair/group constellations 6**

Table 15: Overview of the O developers. See Section 4.2 for information on the data and its representation.

**Sessions OA1 and OA2** Developers O3 and O4 are tasked with writing a test case for some new functionality. Even though they get some help from a colleague along the way, they do not make progress in sessions OA1 and OA2 (separated by a lunch break). There are multiple reasons for this: They neither know that part of the production code nor the underlying technology (React and Redux) nor their development environment so they resort to “console.log”-debugging for which they have to rebuild the software in three-minute cycles. The pair

6. Project homepage: https://impress.js.org

7. Project homepage: https://www.docker.com/

8. Amazon Web Services: https://aws.amazon.com/

speaks English throughout the session, which is neither developer’s first language.

Note on Data: Session OA1 started as plain PP with only one screen. However, O4 started using his laptop along the way (and continued in OA2), but his screen is not recorded. Additionally, there is not webcam recording for session OA1.

Sessions OA3, OA4, and OA5 Four developers of the team work on a bug that causes some rendering issues on drag-and-drop actions in a calendar view. Between OA3 and OA4 there is a 10 minute break; between OA4 and OA5, there is a 15 minute break, during which developers O2 and O4 left the group, leaving just the pair O1/O3 who amend test cases, refactor the production code, and eventually fix the bug. Along the way, O1 explains general software development principles to O3.

Sessions OA6 and OA7 Developers O3 and O4 try to understand a performance issue in their web application for about an hour. Twenty minutes later, O4 continues this task with O1. Both sessions are done via a Skype call with audio and screensharing but no webcam.

Note on Data: No webcam was recorded.

Sessions OA8, OA9 and OA10 The three developers O1, O3, and O4 work on a bug in three sessions. At first, they investigate both production and test code to understand the reason for a newly failing test case that is unrelated to the bug: They changed some implementation but did not adapt the mock objects used in the tests accordingly. They eventually adapt the mock objects and write new test cases to reproduce the bug. Developer O1 leaves session OA8 after 17 minutes, but is again part of the group in sessions OA9 and OA10.

6.13. Company P

Company P develops and operates the web platform for a car part retailer. The website is consists of multiple large PHP apps. During a one-week stay, we recorded 4 sessions with 3 developers. See Table 16 for general information on the developers and Table 29 for structured information on the recorded sessions.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gnd</th>
<th>Lang</th>
<th>Dev. Y</th>
<th>PP Y</th>
<th>Comp. Y</th>
<th>#S #P</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>male</td>
<td>DE</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>4 2</td>
</tr>
<tr>
<td>P2</td>
<td>male</td>
<td>DE</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>2 1</td>
</tr>
<tr>
<td>P3</td>
<td>male</td>
<td>DE</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2 1</td>
</tr>
</tbody>
</table>

Table 16: Overview of the P developers. See Section 4.2 for information on the data and its representation.

Sessions PA1 and PA2 In session PA1, developers P1 and P2 review a database migration written by P1 and later discuss the requirements that led to the database schema change in the first place. They continue after their lunch break with session PA2 where they test and debug the migration and end up changing test cases to remove embedded false assumptions.

Sessions PA3 and PA4 Developers P2 (who is more proficient in the backend) and P3 (more frontend experience) continue the implementation of a new API endpoint which P3 already started. In PA3, P3 shows his existing implementation for which they write tests; P2 explains some backend related software development best practices. On the next day, in session PA4, they continue with implementing the database access which causes them problems because of some idiosyncrasy of their object-relational (OR) mapper. There is a 13-minute break in session PA4 due to a spontaneous team meeting.

7. Discussion

The data collection procedure described in Section 3 has a number of properties which affect the data quality in terms of the kind of data which can be collected this way, but also affecting the developers’ behavior.

7.1. Limitation of Scope

Due to the method’s design, a number of systematic gaps can be expected. That is, there probably exists industrially relevant pair programming behavior which is not captured.

- Not all companies: Due to our naturalistic approach, we did not ask developers to pair program, so we did not target companies with little or no pair programming usage.
- Not all developers: All recordings are voluntary and some developers may not want to be recorded. In company P, for example, one team member was generally inclined to working in pairs, but did not want to be part of the data collection, despite an emphatic recommendation by a colleague who had his first reflective interview.
- No short sessions: The majority of the sessions in the repository is one hour or longer, with the shortest one being about 25 minutes long. In discussions with practitioners, however, some reported of regular session lengths of 10 or 15 minutes. The recording infrastructure is designed for recording individual sessions, not, say, everything that happens on a work station over the course of a day. Since the recording setup poses an overhead to the normal work flow, ad hoc pairings are difficult to record. Furthermore, pairs that had already gone through organizing and setting up a recording session then possibly work longer than they otherwise would have.
- No tense situations: The mere presence of a researcher on site may be regarded as a distraction. In both companies O and P there were multiple months between the first discussion and the start of the main data collection, and in both cases it was due to the Scrum Masters wanting to postpone the research activity until a turbulent phase in the respective projects was over. A second data collection phase in company O did not happen because of immense time pressure for the software developers—even though both developers and Scrum Masters were very happy with the insights from the first round. It is not clear whether any pair programming was done in these stressful phases.
There are other limitations of the data, which are not strictly due to design, but due to practicalities of getting in contact with a company and traveling.

- **Western Cultural Background:** All companies are based in Germany with the exception of company M which is in Oslo, Norway. The L-sessions were individual training sessions and had no organizational background (see Section 6.9). Most developers are native German speakers. The L-developers are the only native English speakers; the M- and O-developers use English as their work language.

### 7.2. Effects of Recording Infrastructure

In companies A to D, the developers’ IDEs were also equipped with a plugin to collect technical information on their current activities, focus, etc. (see [9, pp. 85 & 461–479]). This led to some artifacts in the programming sessions, e.g., in CA2 where the developers spent 1:20 minutes trying (and failing) to look up an ID from the issue tracker on the (remote) development computer before tabbing out of the RDP session to find the necessary information within five seconds on the (local) recording computer which they first avoided to ensure continuous data collection. In session CA3, the IDE was repeatedly unresponsive for 1:20 minutes at a time totaling about one third (!) of the whole session, for which the developer had no explanation and which may be due to a defect in the data collection plugin.

There were several instances of developers not working on their own machine. In sessions CA2 and BA1 the developers spent some time to get comfortable with their IDE as some options are not set as they were used to. Session DA2 starts with several minutes of the pair waiting for an SVN update to complete since the workspace had not been used for a while.

The recording infrastructure was not always fully compatible with the local circumstances and the developers’ habits. In session EA1, the same keyboard shortcut was used for stepwise debugging as for Camtasia to pause recording. Not only did this lead to some gaps in the screencast, but appears to have also slightly confused the developers (as the continuous audio recording reveals). All pairs in company F appear to have used two monitors, but the recording setup at the time was not able to capture both, so the screencast is missing one half.

Although wireless microphones and webcam were set not to bother the developers, they occasionally fiddled with it, e.g., before leaving the desk for a minute and again upon their return. One pair knocked over the webcam from its tripod; another pair took a break together to get some candies and wandered beyond the wireless transmitter radius while still talking about their task.

Reports on how the subjects felt regarding the data collection are available from the C-developers only, some of which say they felt being watched while others claim to have forgotten the camera after five minutes.

### 7.3. Effects of Pre-Existing Notions

We had pre-existing notions of what the social reality of industrial software development looks like which were deeply embedded in the data collection process.

**Not Recording All Aspects** The first such notion is pair programming itself. Our research started with the textbook definition of ‘two developers jointly working on one computer’. Neither of these quantities is fixed in everyday industrial development, but they are in the session recording infrastructure: Screencast software, microphones, and camera angle are all set up for two developers on one computer. However, developers may suddenly open their own laptop or interact with others developers who are out of reach of the microphones and/or beyond the camera angle (as was noted by developer D4 in session DA2 after a third developer joined the pair: “We need more microphones!”).

Another impact from the research interest on the way data was collected can be seen in the sessions recorded by Plonka, who was initially interested in the driver/navigator metaphor. In order to easily see who is control of keyboard and mouse, the camera angle of sessions in companies C, D, E, and F centers on the developers’ hands, but occasionally cuts off their faces. Sessions from the companies focus on the developers’ faces instead.

In both cases, possibly relevant aspects of the PP process are not recorded making the reconstruction more difficult. In the first, there is technical information missing, in the second, the developers faces and viewing direction makes interpreting their behavior more difficult at times.

**Affecting Developer Behavior** The second, related notion is that of a session pertaining to a task as a meaningful unit the software developers’ workday. Some companies form pairs independent of concrete tasks for multiple days on end during which the pair members behave as one, taking coffee breaks together without really starting or ending a “session”. The data collection procedure described above is session-centric. Questionnaires before and/or after the recording frame the session in two senses. First, they set a ceremonial start and end: In the beginning of sessions CA2 and EA1, the developers filling out the questionnaire was accidentally recorded—it took them more than nine minutes, which might be an unnatural break. Second, the pre-session questionnaire asked the developers to think about the work time ahead. In particular, the questionnaire asked for task classification and description, a characterization of the expected difficulties, and an estimated time to completion (see Fig. 2). Although this may yield valuable context information for the researcher, it may impose an unnatural focus on the developers by making them think about aspects they would not have thought about had it not been for the recording.

Another effect of session-centrism can be observed in multiple session recordings in various companies: Pair members are occasionally interrupted by their colleagues with technical or organizational concerns. A common reaction of the pairs in the recordings is to send away the interrupter unsatisfied (as seen in session AA1 at 1:47:05) or to point out that they just walked into a recording (as seen in session KA1 at 0:04:30), as if the pair programmers wanted to protect the integrity of
the data collection.\textsuperscript{10} The recordings in company $E$ are peculiar in another way, which also possibly indicates an intention of the developers to protect the data collection. The work station for the session recordings was set up in a meeting room, such that the pairs worked completely secluded from the rest of their team. Apparently this was the company’s decision, but it seems unlikely that this reflects the team’s normal pair programming practice.

Such effects are more pronounced in the recordings that were done under a “Understand PP” headline (see Table 1). For instance, the pairings in companies $C$ to $F$ do not appear to be holding up to the naturalistic ideal as only one of the overall 21 pairs was recorded twice.\textsuperscript{11} In the spirit of a “workshop”, the developers could write their names on a list with a morning and an afternoon slot if they wanted to be recorded (see also Appendix B).

In company $P$, the team at one point discussed how the next pair should be formed based on what we at the time understood as organizational constraints. However, in the next reflective interview, the developers revealed to us that they intended to give another, previously not-recorded colleague the chance to also benefit from the feedback we could provide. Again, the employed data collection method did interfere with a “natural” formation of pairs.

7.4. Summary of Data Quality

Notwithstanding the above limitations of the data collection—some a direct result of the way the involved researchers collected data (limited breadth of data collection which complicates interpretation, also: framing), others resulting from reactions of the subjects to the researchers’ presence (protecting integrity, longer sessions)—the session repository still comprises diverse, realistic, detailed data. At the time of writing, it contains 67 recordings from 13 different companies featuring 57 different professional software developers who worked together in 41 different constellations of (mostly) two members.

In all of these sessions, the developers worked on their normal tasks for as long as they wanted, and in most cases, the developers also freely chose who to work with and when to start. The exception here are the 23 sessions from companies $C$ to $E$, where the developers had to sign up for either a morning or an afternoon slot, and were possibly inclined to work with partners they would normally not pair with in the prospect of learning something in the reflective interview. It can also not be ruled out that the developers in all companies generally felt compelled to work in pairs more often than normal for the same reason.

All of the above concerns may affect the frequency of phenomena (such as more or less conflict situations, more or less easy tasks, or more or less fatigue due to longer sessions), but none of these appear to produce to entirely artificial behavior. Depending on the make up of concrete studies (e.g., qualitative or quantitative approach), the above considerations need to be kept in mind.

8. Usage in Publications

See Figure 4 for an overview of all publications from our research group that build on the data of the PP-ind repository described in this document.

References

Figure 4: Timeline of data collection and scientific publications originating in our research group. Arrows between publications indicate reuse of ideas or building on results; arrows from data collection indicate which PP sessions were analyzed. PhD theses are set bold.

Table 17: Overview of the A sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA1</td>
<td>PP</td>
<td>A1 A2</td>
<td>2007-01-26 13:43</td>
<td>2:22</td>
<td>DE</td>
<td>Java, Objective-C, SQL, HTML, Tcl</td>
<td>2 1</td>
<td>2 1</td>
</tr>
</tbody>
</table>

Table 18: Overview of the B sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA1</td>
<td>PP</td>
<td>X X</td>
<td>2007-09-14 13:38</td>
<td>1:47</td>
<td>DE</td>
<td>PHP</td>
<td>1- 2</td>
<td>2 2</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA1</td>
<td>PP</td>
<td>X X X</td>
<td>2008-05-05 13:27</td>
<td>1:18</td>
<td>DE</td>
<td>Java</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>CB1</td>
<td>PP</td>
<td>X</td>
<td>2008-05-06 07:54</td>
<td>1:12</td>
<td>DE</td>
<td>Java</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CA2</td>
<td>PP</td>
<td>X X X</td>
<td>2008-05-07 11:46</td>
<td>1:14</td>
<td>DE</td>
<td>Java</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>CA3</td>
<td>PP</td>
<td>X X X</td>
<td>2008-05-07 15:34</td>
<td>2:10</td>
<td>DE</td>
<td>Java</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CA4</td>
<td>PP</td>
<td>X X</td>
<td>2008-05-08 10:25</td>
<td>1:34</td>
<td>DE</td>
<td>Java</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CA5</td>
<td>PP</td>
<td>X X</td>
<td>2008-05-09 10:32</td>
<td>1:23</td>
<td>DE</td>
<td>Java</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 19: Overview of the C sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA1</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-06 14:07</td>
<td>2:21</td>
<td>DE</td>
<td>Java, XML</td>
<td>“3–4”</td>
<td>1</td>
</tr>
<tr>
<td>DA2</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-08 10:13</td>
<td>2:23</td>
<td>DE</td>
<td>Java</td>
<td>“first time”</td>
<td>2</td>
</tr>
<tr>
<td>DA3</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-08 14:01</td>
<td>1:05</td>
<td>DE</td>
<td>XML (Spring)</td>
<td>6</td>
<td>2 5</td>
</tr>
<tr>
<td>DA4</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-08 16:22</td>
<td>2:00</td>
<td>DE</td>
<td>Java, XML</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>DA5</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-09 10:27</td>
<td>0:31</td>
<td>DE</td>
<td>Java</td>
<td>1</td>
<td>n.a.</td>
</tr>
<tr>
<td>DA6</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-09 13:01</td>
<td>0:58</td>
<td>DE</td>
<td>Java</td>
<td>n.a.</td>
<td>“1 / 4”</td>
</tr>
</tbody>
</table>

Table 20: Overview of the D sessions. See Section 4.3 for information on the data and its representation. All pairs filled out the pre-session questionnaire together. In session DA2, the pair talks with two other developer (5 minutes with D7, then 12 minutes with D6). In session DA6, the pair worked on two somewhat separate tasks which D2 rated separately in the post-session questionnaire.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA1</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-27 11:29</td>
<td>1:17</td>
<td>DE</td>
<td>C++</td>
<td>“2–3”</td>
<td>3</td>
</tr>
<tr>
<td>EA2</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-27 13:18</td>
<td>2:46</td>
<td>DE</td>
<td>XML (Maven)</td>
<td>“first time”</td>
<td>“–”</td>
</tr>
<tr>
<td>EA3</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-28 10:43</td>
<td>1:25</td>
<td>DE</td>
<td>C++</td>
<td>3</td>
<td>3 3</td>
</tr>
<tr>
<td>EA4</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-29 09:37</td>
<td>1:52</td>
<td>DE</td>
<td>C#</td>
<td>“–”</td>
<td>“0”</td>
</tr>
<tr>
<td>EA5</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-29 13:09</td>
<td>1:41</td>
<td>DE</td>
<td>C++, Java</td>
<td>“none”</td>
<td>“5–6”</td>
</tr>
<tr>
<td>EA6</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-30 10:05</td>
<td>1:40</td>
<td>DE</td>
<td>C++</td>
<td>“2–3”</td>
<td>3</td>
</tr>
<tr>
<td>EA7</td>
<td>PP</td>
<td>X X</td>
<td>2008-10-31 09:25</td>
<td>1:50</td>
<td>DE</td>
<td>C++</td>
<td>2 2</td>
<td>2 2</td>
</tr>
</tbody>
</table>

Table 21: Overview of the E sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA1</td>
<td>PP</td>
<td>X X</td>
<td>2008-11-11 15:06</td>
<td>1:45</td>
<td>DE</td>
<td>Java, XML, SQL</td>
<td>3</td>
<td>3 2 2</td>
</tr>
<tr>
<td>FA2</td>
<td>PP</td>
<td>X X</td>
<td>2008-11-12 11:07</td>
<td>1:59</td>
<td>DE</td>
<td>Java, XML, SQL</td>
<td>3</td>
<td>3 1 2</td>
</tr>
<tr>
<td>FA3</td>
<td>PP</td>
<td>X X</td>
<td>2008-11-13 11:06</td>
<td>2:39</td>
<td>DE</td>
<td>Java</td>
<td>2</td>
<td>3 2 1</td>
</tr>
<tr>
<td>FA4</td>
<td>PP</td>
<td>X X</td>
<td>2008-11-13 15:18</td>
<td>1:52</td>
<td>DE</td>
<td>Java</td>
<td>2</td>
<td>2 3 1</td>
</tr>
</tbody>
</table>

Table 22: Overview of the F sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>JA1</td>
<td>DPP</td>
<td>X X</td>
<td>2013-01-31 14:05</td>
<td>1:07</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA2</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-13 10:51</td>
<td>1:15</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA3</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-13 13:17</td>
<td>1:53</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA4</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-13 15:26</td>
<td>2:01</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA5</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-14 10:33</td>
<td>1:36</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA6</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-14 13:13</td>
<td>0:42</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA7</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-14 14:54</td>
<td>1:56</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA8</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-15 10:54</td>
<td>1:06</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>JA9</td>
<td>DPP</td>
<td>X X</td>
<td>2013-02-15 12:59</td>
<td>5:27</td>
<td>DE</td>
<td>Java</td>
</tr>
</tbody>
</table>

Table 23: Overview of the J sessions. See Section 4.3 for information on the data and its representation.
<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>KA1</td>
<td>PP</td>
<td>X X</td>
<td>2013-03-14 10:37</td>
<td>2:00</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>KA2</td>
<td>PP</td>
<td>X X</td>
<td>2013-03-14 13:15</td>
<td>2:53</td>
<td>DE</td>
<td>Java</td>
</tr>
<tr>
<td>KB1</td>
<td>PP</td>
<td>X X</td>
<td>2013-05-02 13:45</td>
<td>0:53</td>
<td>DE</td>
<td>Java, SQL</td>
</tr>
<tr>
<td>KB2</td>
<td>PP</td>
<td>X X</td>
<td>2013-05-02 15:26</td>
<td>1:36</td>
<td>DE</td>
<td>Java, SQL</td>
</tr>
<tr>
<td>KC1</td>
<td>PP</td>
<td>X X</td>
<td>2013-10-29 11:24</td>
<td>0:59</td>
<td>DE</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>KC2</td>
<td>PP</td>
<td>X X</td>
<td>2013-10-29 12:59</td>
<td>2:01</td>
<td>DE</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>KC3</td>
<td>PP</td>
<td>X X</td>
<td>2013-11-29 11:00</td>
<td>0:53</td>
<td>DE</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>KC4</td>
<td>PP</td>
<td>X X</td>
<td>2013-11-29 11:44</td>
<td>2:10</td>
<td>DE</td>
<td>CoffeeScript</td>
</tr>
</tbody>
</table>

Table 24: Overview of the K sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA1</td>
<td>RPP</td>
<td>X X</td>
<td>2014-02-27 20:41</td>
<td>1:00</td>
<td>EN</td>
<td>Ruby</td>
</tr>
<tr>
<td>LB1</td>
<td>DPP</td>
<td>X X</td>
<td>2014-03-06 16:11</td>
<td>0:47</td>
<td>EN</td>
<td>HTML, CSS</td>
</tr>
</tbody>
</table>

Table 25: Overview of the L sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA1</td>
<td>PP</td>
<td>X X</td>
<td>2014-10-16 11:42</td>
<td>0:25</td>
<td>EN</td>
<td>SQL</td>
</tr>
</tbody>
</table>

Table 26: Overview of the M sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA1</td>
<td>SbS</td>
<td>X X</td>
<td>2016-01-12 13:30</td>
<td>0:47</td>
<td>DE</td>
<td>–</td>
</tr>
<tr>
<td>NA2</td>
<td>SbS</td>
<td>X X</td>
<td>2016-01-12 13:37</td>
<td>0:41</td>
<td>DE</td>
<td>–</td>
</tr>
<tr>
<td>NA3</td>
<td>SbS</td>
<td>X X</td>
<td>2016-01-12 15:06</td>
<td>0:57</td>
<td>DE</td>
<td>–</td>
</tr>
<tr>
<td>NA4</td>
<td>SbS</td>
<td>X X</td>
<td>2016-01-12 15:12</td>
<td>2:04</td>
<td>DE</td>
<td>–</td>
</tr>
<tr>
<td>NA5</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-01-18 12:46</td>
<td>3:29</td>
<td>DE</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 27: Overview of the N sessions. See Section 4.3 for information on the data and its representation.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA1</td>
<td>PP</td>
<td>X X</td>
<td>2016-06-01 10:51</td>
<td>1:24</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA2</td>
<td>PP</td>
<td>X X X</td>
<td>2016-06-01 13:27</td>
<td>1:32</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA3</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-06-08 14:55</td>
<td>0:59</td>
<td>DE, EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA4</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-06-08 16:05</td>
<td>0:50</td>
<td>DE, EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA5</td>
<td>PP</td>
<td>X X</td>
<td>2016-06-08 17:11</td>
<td>1:09</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA6</td>
<td>RPP</td>
<td>X X</td>
<td>2016-06-09 14:00</td>
<td>0:56</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA7</td>
<td>RPP</td>
<td>X X</td>
<td>2016-06-09 15:17</td>
<td>1:39</td>
<td>DE</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA8</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-06-15 13:47</td>
<td>1:16</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA9</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-06-15 15:16</td>
<td>0:47</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
<tr>
<td>OA10</td>
<td>Mob</td>
<td>X X X</td>
<td>2016-06-15 16:50</td>
<td>1:44</td>
<td>EN</td>
<td>CoffeeScript</td>
</tr>
</tbody>
</table>

Table 28: Overview of the O sessions. See Section 4.3 for information on the data and its representation. Session OA2 started as a PP (with Active Observer) session, and developer O6 joined for a period of 14 minutes. Session OA8 started as a Mob Programming session, but O1 had to leave for a meeting after 17 minutes.

<table>
<thead>
<tr>
<th>ID</th>
<th>Mode</th>
<th>Developer</th>
<th>Start</th>
<th>Dur.</th>
<th>SL</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA1</td>
<td>PP</td>
<td>X X</td>
<td>2018-06-05 11:24</td>
<td>0:58</td>
<td>DE</td>
<td>PHP</td>
</tr>
<tr>
<td>PA2</td>
<td>PP</td>
<td>X X</td>
<td>2018-06-05 13:35</td>
<td>1:30</td>
<td>DE</td>
<td>PHP</td>
</tr>
<tr>
<td>PA3</td>
<td>PP</td>
<td>X X</td>
<td>2018-06-06 12:23</td>
<td>1:31</td>
<td>DE</td>
<td>PHP</td>
</tr>
<tr>
<td>PA4</td>
<td>PP</td>
<td>X X</td>
<td>2018-06-07 11:09</td>
<td>1:42</td>
<td>DE</td>
<td>PHP</td>
</tr>
</tbody>
</table>

Table 29: Overview of the P sessions. See Section 4.3 for information on the data and its representation.
Appendix A. Recording Technicalities

We used three different generations of recording setups, the first of which was developed by Laura Plonka and Stephan Salinger, the other two by Julia Schenk and Franz Zieris.

Generation 1

Two computers are used in this setup: The development environment of the developers runs on one machine whose output is mirrored to another computer where the screen content is recorded with TechSmith Camtasia. On Linux machines (company A), developers would work locally while the screen content is transferred to a Windows machine via VNC (Virtual Network Computing) to be recorded remotely (see Fig. 5b). On Windows machines (companies C, D, E, F), the developers would sit on the recording machine and use RDP (Remote Desktop Protocol) to log on the development machine, such that the recording device only needs to run Camtasia and the RDP client (see Fig. 5c). In some cases (company B), the same machine would be used for both development and recording (see Fig. 5a) thus limiting the available resources available in the developers’ environment.

For the audio channel, there were also two solutions. The easy one was to use the webcam microphone (companies A and B, Figs. 5a and 5b), which was not able to pick up the developers’ speech independently making it difficult to understand them when both speak at the same time. For higher audio quality, two wireless microphones were used in later recordings (companies C to F, Fig. 5c). Using an external USB soundcard, the two mono channels were hard-panned (100% left and 100% right) and mixed to one stereo channel.

Generation 2

For the recordings in companies J and K, we used the web conferencing tool Adobe Connect which relies on a Flash-based browser plugin to share one’s screen content, webcam, and microphone.

To record the distributed pair programming sessions of company J, two Adobe connect meetings were started such that both developers each shared their respective screen, webcam, and headset microphone (which they had to use anyway for their DPP session). On the receiving end, the researcher stacks the two screen outputs on top of each other on a vertically rotated monitor and records the whole configuration (see Fig. 5d). Schenk organized the setup and the recording of the J-sessions; see [13, Sec. 3.2.2 & 3.2.3] for more details.

In companies K and M a similar setup was used: The developers put the headbands of USB headsets in the nape of their necks such that the cushions do not cover their ears and the microphone can pick up their speech. Again, two Adobe Connect meetings were necessary, this time to transmit both microphone signals. In company K, the pairs’ dual-screen setup could also be captured this way (see Fig. 5e).

Compared to the first generation, the advantages of this infrastructure were (a) that the session could be watched live by the researcher while it was still recording thus making the next step (quick analysis) a bit faster and (b) the ease of the setup allowed that the recording could be done completely remotely as the developers had all necessary equipment on-site (modern web-browser, webcam, and headsets). Disadvantages were that (c) this setup only worked for Windows machines and that support for Flash (necessary for running Adobe Connect) was already declining at the time, and (d) fluctuation in network latency for the two audio channels meant that, for colocated pair programming in companies K and M, there was a notable randomly changing offset between the two signals. Whenever both microphones picked up the sound of one developer, e.g., because they looked at each other, an annoying robotic echo could be the result, which makes listening to the record rather unpleasant.

Generation 3

In order to enable recordings of pair programming sessions independent of the developers’ operating systems, we looked for other solutions than Adobe connect. The most versatile ones turned out to be Skype and TeamViewer, the latter of which was able to transmit and record pixel-accurate screen contents so our research group purchased a commercial license. TeamViewer’s recording feature allows to losslessly record the contents of a, say, Full HD display even if the recording device has a much smaller resolution available such as a window on a Laptop screen (see Fig. 5f). In order to process TeamViewer data further, however, the recording first needed to be rendered as a pixel video, which at times took much longer than the recording itself.

There were two options for recording the webcam. For relatively small distances between the development and the recording computers, the webcam could be connected directly to the recording machine using a really long USB cable in order to not disturb the developers. For setups more than one room apart, the webcam feed was transmitted in the TeamViewer session, which posed additional problems: Even though the webcam video is included in the pixel-video export, it has a fixed position and small resolution (ca. 160x120 pixels, next to the screen cast, in the upper right corner), and is also often notably out of sync with the screencast. To compensate this, we used Camtasia to screen capture the webcam feed separately while the session is running at a resolution of about 460x340 (see schematic in Fig. 5f, and annotated photograph from recording on-site in Fig. 6).

To avoid the robotic echo of network-transmitted dual audio, we record the audio of co-located sessions locally without any network in between. In company O, we used the same wireless microphones as before and fed them into one dictaphone; in company P we used two independent

15. Product homepage: https://www.skype.com/
Figure 5: Different recording setups. Developers sit in front of a screen with a webcam and some means to record audio. Screen contents are transferred to another machine (together with audio and/or webcam signal), where they are recorded with a screen capture tool.

Figure 6: Concurrent recording of screencast and webcam as shown in Fig. 5f. This is an annotated photograph taken during the recording of session PA2 in a conference room at company P. The developers P1 and P2 sit two rooms further down the hall. We used an external monitor only to follow the session live; it would not have been strictly necessary because TeamViewer will record the screencast in full resolution regardless of its current display size.
dictaphones each wired to a simple lapel microphone to record the developers individually.  

**Synchronizing the Data Channels**

Depending on the setup, a recording session would yield a number of files with a number of data streams. In the simplest form (as for companies A and B) the result would be a single Camtasia file containing the screencast stream and the webcam stream (which also includes the audio). Such a file could be opened directly in Camtasia Studio and exported to a self-contained video file such as an AVI or MP4 video file.

In companies C through F—where audio was recorded externally—there were two Camtasia files: One comprising the screen content and webcam video just as in the simplest setup; the other containing the externally recorded dual-mono-mixed-to-stereo audio signal. To ease synchronizing the separate audio and video channels, the developers in these recordings were asked to clap in front of the camera. Ultimately, there were two ‘timelines’ that needed to be aligned.

With the second generation recording infrastructure using Adobe Connect web conferences, no such alignment was strictly necessary as all individual streams (two screens, one or two webcams, two audio channels) were already synchronized the moment they were captured off the researcher’s screen. However, some audio post-processing was still necessary as Adobe Connect center-panned both audio channels during the meeting such that they were mixed together in the Camtasia recording. Luckily, Adobe Connect allowed to record the sessions on the conference server and provided a large ZIP file containing all individual streams, including the audio track as an MP3 file. Using the mixed audio from the Camtasia recording as a reference, we could align the two separate audio recordings and hard-pan them to ultimately get a self-contained video after aligning these three timelines.

In the third recording generation, which relies on TeamViewer, a total of four timelines need to be synchronized: The screencast; two independent dictaphone recordings which, even though they come from the same model with identical settings, differ inexplicably in recording speed by several seconds per hour; and the webcam feed. To synchronize these, we first stretched one of the audio recordings (without changing the pitch) to match the other, then looked for isolated code changes in the screencast and listen for audible keystrokes, then noticed that the either the screencast or audio needed to be stretched again as they, too, drifted further and further apart towards the end of the session, and finally did the same again for the webcam video (looking for bilabial plosives in the developers’ speech, such as a [p] or [b]) which sometimes also needed some stretching by yet a different factor.

For a reproducible production from the raw material involving the various synchronization steps, we use the open-source video scripting system AviSynth.  

An AviSynth script is executed by the AviSynth frameserver which then serves compatible programs such as VirtualDub with video frames and audio samples. These can either be played back in realtime (at least for simple scripts) or be sent to an encoder such as x264 to obtain a self-contained video file. Apart from being reproducible, video production through AviSynth did not put constraints on the video dimensions (that is, no constraints on top of the H.264 standard)—unlike Camtasia Studio which up to and including version 8 only supported a maximum resolution of 2048 by 2048 pixels.

Overall, the production of a single session video in the last setup would take at least one day of work with many iterations of manual finetuning.

---

17. Dictaphone model: Olympus VN-8700PC; Microphone model: AV-JEFET TCM 141
19. Homepage: http://www.virtualdub.org/
20. Homepage: https://www.videolan.org/developers/x264.html
WORKSHOP PAAR PROGRAMMIERUNG

BERICHT

WORKSHOP PAAR PROGRAMMIERUNG

bei

05.05-09.05. 2008

Laura Plonka
Freie Universität Berlin
AG Software Engineering
1 Einleitung und Motivation ................................................................. 3
2 Durchführung des PP-Workshop ...................................................... 3
  2.1 Vorbereitung ................................................................................. 3
  2.2 Einführungsveranstaltung .............................................................. 4
  2.3 Datenaufzeichnung ....................................................................... 4
  2.4 Analyse der Sitzungen ohne die Programmierer ......................... 5
  2.5 Reflexions- und Feedback-Runde mit den Programmieren ......... 6
  2.6 Abschluss-Runde mit allen Programmieren ............................... 7
3 Ergebnisse des Workshops ............................................................... 7
  3.1 Nutzen der Paar Programmierung bei .......................................... 7
  3.2 Erkenntnisse aus den Feedback-Runden ..................................... 8
  3.3 Erkenntnisse aus der Abschluss-Runde ...................................... 11
4 Zusammenfassung und Ausblick ..................................................... 12
5 Referenzen ....................................................................................... 13
1 Einleitung und Motivation


Während der Paar Programmierung (PP) arbeiten die Entwickler meist sehr konzentriert. Dabei bleibt wenig Möglichkeit den eigenen PP-Prozess bewusst wahrzunehmen und zu reflektieren. Also wie kann man Paar Programmierer individuell in der Weiterentwicklung ihres PP-Prozesses unterstützen?


Hierfür zeichnen wir die PP-Sitzung eines Paars in ihrem Arbeitsumfeld auf, analysieren die Daten und diskutieren die Ergebnisse am nächsten Tag anhand des aufgezeichneten Materials (ca. eine halbe Stunde lang) mit dem Paar.

Themen: Was lief gut? Was lief nicht gut? Warum?
Was könnte man am Verhalten ändern wollen?

2 Durchführung des PP-Workshop

Im folgenden Abschnitt wird die Vorbereitung sowie der Ablauf des Workshops beschrieben.

2.1 Vorbereitung

Vor dem Beginn des Workshops wurden die folgenden Punkte bearbeitet:

- Einrichten der Infrastruktur


Laura Plonka – Freie Universität Berlin

Seite 3
Entwicklungsrechner

Remote-Desktop
Verbindung

Aufzeichnungsrechner

Arbeitsplatz für Paare während des Workshops
Installation von
- Camtasia
und
- Webcam

Entwicklungsrechner
Installation von
- ECG

Abbildung 1: Schematische Darstellung der Infrastruktur, die Entwickler saßen in der Workshop-Woche, wenn sie sich aufzeichnen lassen wollten am Aufzeichnungsrechner, arbeiteten aber wie gewohnt auf ihrem Entwicklungsrechner.

- **Entwurf Geheimhaltungsvereinbarung und Einverständniserklärung**

Um den Umgang mit den erhobenen Daten zu regeln, wurde von einer Geheimhaltungserklärung entworfen. Die FUB entwickelte ihrerseits eine Einverständniserklärung für die einzelnen Paar Programmiere, die sicherstellt, dass sie Programmierer wissen, welche Daten erhoben werden und freiwillig an der Aufzeichnung teilnehmen.

### 2.2 Einführungsveranstaltung

Zu Beginn des Workshops wurde eine Einführungsveranstaltung von der FUB für die Entwickler durchgeführt. Ziel war, die Entwickler über den Nutzen und die Motivation, sowie den Ablauf zu informieren. Zusätzlich wurde vorgestellt, welche Daten von den Entwicklern erhoben werden. Diese Einführungsveranstaltung dauerte ca. 1 Stunde.

### 2.3 Datenaufzeichnung


Die Aufzeichnungsdauer pro Session wurde auf 1,5-2 Stunden beschränkt, es gab aber weder Einschränkungen bzgl. der Paarkonstellationen noch der Aufgaben, so dass die Paare während der Aufzeichnung ihren normalen Programmieraufgaben nachgehen konnten.

Während der Aufzeichnung wurden die folgenden Daten erhoben:

- Video der Programmierer mittels Webcam
- Audio der Programmierer über wireless Mikrophone
- Desktop der Programmierer
- Computer Log-Files (Loggen der Anwendung im Fokus, Dateiname, Anzahl der Editier-Schritte)


Zusätzlich wurden vor dem Start und nach dem Ende der Aufzeichnung von den Paaren jeweils einen kurzen Fragebogen (ca. 10 Min je Fragebogen) ausgefüllt.
2.4 Analyse der Sitzungen ohne die Programmierer

Im ersten Schritt der Analyse wurden die Daten für die Reflexions- und Feedback-Runde aufbereitet und von der FUB analysiert.

- Automatisierte Auswertung

- Manuelle Videoanalyse
  Das Video (Desktop und Videobild der Programmierer) wurde zusätzlich noch manuell am Tag der Aufzeichnung von der FUB analysiert (Aufwand: ca. 2x Dauer des Videos). Dabei wurden zum Beispiel Informationen über das Driver-Observer-Verhalten sowohl quantitativ als auch qualitativ ausgewertet.


Abbildung 2 zeigt den Prozess der Analyse mittels *WhoTalks*. Die resultierende Visualisierung (siehe Abbildung 3) und die Erkenntnisse aus der Analyse dienen als Einstieg für die Reflexions- und Feedbacksitzung.

Abbildung 2: Schritte der Analyse; Im ersten Schritt werden die Daten quantitativ ausgewertet, die Evaluation ermöglicht einen statistischen Überblick und eine Visualisierung. Im zweiten Schritt werden die Videos qualitative analysiert und auffällige Muster bzw. Verhaltensweisen identifiziert und mit den Informationen aus den Fragebögen in Beziehung gesetzt und visualisiert.
2.5 Reflexions- und Feedback-Runde mit den Programmieren

Die Reflexions- und Feedback-Runde fand jeweils am Tag nach der Aufzeichnung statt (mit Ausnahme des Freitags).

1. Fragen- und visualisierungsgestützte Reflexion

Anhand einer Liste von Fragen und durch die Visualisierung der Sitzung konnten die Paare den Verlauf der Sitzung im ersten Schritt nochmals nachvollziehen. Unter anderem wurden die folgenden Fragen von den Entwicklern diskutiert:

- Wie verlief die Sitzung?
- War etwas störend?
- Was war besonders hilfreich bei der Zusammenarbeit?
- Wie gut schätzen sie ihre Zusammenarbeit ein?
- Warum war es sinnvoll diese Aufgabe im Paar zu lösen?
- Wie schätzen Sie Ihre Driver-Anteile ein?

In diesem Rahmen konnten Probleme, wie auch positive Verhaltensweisen von dem Paar erörtert werden. Häufig wurden im Zuge dieser Reflexion Bereiche, die bereits während der vorab Video-Analyse aufgefallen waren.

2. Vorstellung der Ergebnisse der Videoanalyse: Externe Anregung und Feedback

Im zweiten Schritt erfolgte die Diskussion der Ergebnisse aus der Videoanalyse. Falls die Entwickler im ersten Schritt bereits die Aspekte thematisierten wurden diese mit dem externen Eindruck aus der Videoanalyse in Beziehung gesetzt und auch die externe Sicht diskutiert. Falls nicht alle Aspekte angesprochen wurden, die in der Video-Analyse identifiziert werden konnten, wurden die auffälligen Verhaltensweisen vorgestellt und mit dem Paar diskutiert.

Speziell wenn die Paare problematische Verhaltensweisen ansprachen, wurden Hinweise und Muster vorgestellt wie andere Paare solche Probleme lösen. In dem Abschnitt „Ergebnisse des Workshops“ sind die konkreten Aspekte beschrieben, die im Rahmen dieser Sitzungen diskutiert wurden. Insgesamt dauerte die Reflexions- und Feedback-Runde ca. eine halbe Stunde pro Paar.
2.6 Abschluss-Runde mit allen Programmieren
Zum Ende der Workshop-Woche wurde eine Abschluss-Runde mit allen Programmieren durchgeführt, um die Woche zu reflektieren. Die Themen waren:

- Vorstellung der Erkenntnisse aus dem Workshop
- Was hat der Workshop bzw. den einzelnen Entwicklern gebracht?
- Was kann an dem Reflexionsprozess (organisatorisch wie auch inhaltlich) verbessert werden?
- Was passiert im Anschluss des Workshops mit den Daten?

Die Ergebnisse der Abschluss-Runde sind in Abschnitt „Erkenntnisse aus der Abschluss-Runde“ diskutiert.

3 Ergebnisse des Workshops

Alle Programmierer (7 Entwickler und Entwicklungsleitung) aus dem Team nahmen mindestens einmal an der Aufzeichnung teil.

Die Auswertungen der vorab Fragebögen zeigt, dass die Paare ihre Zusammenarbeit als unterschiedlich gut einschätzten; vier der sechs Paare schätzten ihre Zusammenarbeit als gut bis sehr gut ein und zwei Paare als befriedigend bis ausreichend.

3.1 Nutzen der Paar Programmierung bei PP
Es gibt eine Reihe von Studien in der Wissenschaft, die den Nutzen von Paar Programmierung untersuchen. Diese Ergebnisse sind aber je nach Unternehmensgröße und Motivation für den Einsatz von PP nicht zwangsläufig übertragbar. Daher wurde im Rahmen der Reflexions- und Feedbackrunde mit den Entwicklern besprochen, warum sie PP bei PP einsetzen, mit folgendem Ergebnis:

- **Breitere Wissensbasis**

- **Wissensaustausch**

- **Mehr Disziplin**
  Speziell bei Aufgaben, die sehr viel Disziplin verlangen, wie beispielsweise das Schreiben von Tests unterstützen sich die Programmierer gegenseitig dabei, die Aufgabe zu erledigen. Die Partner ermutigen, in dem sie sich entweder aktiv auffordern die Aufgabe zu Ende zu führen oder aber einer der Partner einfach weiter daran arbeitet und somit der andere eine positive Art des Gruppenzwangs spürt.

Laura Plonka – Freie Universität Berlin
**Bessere Konzentration**

Die Zusammenarbeit fördert die Konzentration und dadurch auch einen besseren Flow. Dadurch wird auch das gemeinsame Lernen (wenn sich beide bisher mit einem bestimmten Themengebiet nicht auskennen) vereinfacht.

**Mehr Spaß**

Die Entwickler haben mehr Spaß an ihrer Arbeit und sind zufriedener.

### 3.2 Erkenntnisse aus den Feedback-Runden

Im folgenden Abschnitt werden die Themen und Erkenntnisse aus den Feedback-Runden anonymisiert vorgestellt, so dass kein Rückschluss auf einzelne Programmierer möglich ist. Die Beschreibung und Diskussion der Themen erfolgt nicht quantitativ, d.h., die Ergebnisse werden nicht statistisch ausgewertet oder die Häufigkeit des Auftretens einzelner Aspekte gezählt, sondern die Aspekte werden qualitativ beschrieben.

#### 3.2.1 Paarspezifische Aspekte

Im Folgenden werden Aspekte vorgestellt, die im Rahmen der Feedback-Sitzung von den Entwicklern diskutiert wurden und die den Paar Programmierungsprozess direkt betreffen.

**Driver-Observer Verhalten**

Allein die Driver-Observer-Verteilung (wer war wie lange Driver) lässt keinen Aufschluss auf Zusammenarbeit im Paar zu. So gab es Paare, die mit der Rollenverteilung zufrieden waren, die eine sehr ausgeglichene Rollenverteilung hatten als auch Paare, bei denen ein Partner den Großteil der Sitzung als Driver arbeitete.

Aber die Kombination aus der Rollenverteilung und der Art des Wechselverhaltens (Wechsel der Rollen Driver und Observer) war ein entscheidender Faktor. Bei Paaren, die mit der Rollenverteilung nicht zufrieden waren und nicht gut eingespielt waren (Information aus dem Fragebogen), verlief das Wechselverhalten häufig ohne Kommunikation. Dabei war einer der Partner meist sehr dominant und griff sich die Tastatur und Maus sehr schnell. Der weniger dominante Partner versuchte zwar häufig die Mause zu greifen, war aber immer nur sehr kurz Driver, bevor der dominante Partner wieder aktiv nach der Tastatur und Maus griff. Auffallend war auch, dass der dominante Partner die Maus und Tastatur auch dann nicht losließ, wenn er sie gerade nicht verwendete (eine Art Reservierungsverhalten).

Sehr gut eingespielte Paare (Information aus dem Fragebogen) Paaren waren zufrieden mit ihrer Rollenverteilung. Bei ihnen verlief der Wechsel in einigen Fällen ebenfalls ohne Kommunikation, aber in gegenseitigem Einverständniss (häufig nachdem eine Teilaufgabe erledigt wurde), in anderen Fällen gab es einen Wechselprozess, in dem sich die Partner aktiv über den Wechsel absprachen, z.B. „Möchtest Du das machen?“, „Soll ich mal wieder?“.

**Pausen**

Während der PP-Sitzungen arbeiten die Entwickler sehr konzentriert und fokussiert. Obwohl immer zwei Programmierer und somit zwei verschiedene Sichten auf das Problem vorhanden sind, kann diese Fokussiertheit auch dazu führen, dass das Paar gemeinsam an einem (eventuell suboptimalen) Ansatz festhält. Eine Pause kann dazu führen, dass das Paar die Aufgabe noch einmal auf einer konzeptionellener Ebene betrachtet und sich später für einen alternativen Ansatz entscheidet.
• **Arbeitsplatz und Rechnerwahl**

Im Rahmen des Workshops trat der Aspekt der Arbeitsplatzwahl nicht auf. Dennoch wurde von den Entwicklern auch angesprochen, dass die Arbeitsplatzwahl bzw. die Rechnerwahl Einfluss auf die PP haben kann. Denn jeder Entwickler konfiguriert seinen Rechner und vor allem die Entwicklungsumgebung individuell. Dies kann dazu führen, dass der Partner, an dessen Rechner nicht gearbeitete wird, sich eventuell nicht so gut zurechtfindet oder er beginnt die Konfigurationen anzupassen und somit die Einstellungen des anderen zu verändern.

• **Blickwinkel auf den Monitor**


Als eines der wichtigsten Themen wurde das Kommunikationsverhalten zwischen den Entwicklern diskutiert. Eine ausgewogene Kommunikation in Form eines Dialoges der Partner ist für den oben genannten Nutzen von PP Voraussetzung. Die statistische Auswertung der automatisierten Sprachanalyse (in [3]) zeigt, dass die Kommunikationsanteile (% der gesprochenen Zeit in Relation zur Gesamtänge der Sitzung für die einzelnen Programmierer) bei allen Paaren ziemlich ausgeglichen waren. Folgende Punkte wurden im Bezug auf die Kommunikation genauer diskutiert:

• **Entscheidungsfindungsprozess**


Eine andere Lösung ist, wenn der Observer aktiv nachfragt, sobald er merkt, dass er den Lösungsansatz nicht mehr nachvollziehen kann. Auch dieses aktive Nachfragen des Observers trat innerhalb von PP-Sitzungen bei auf.
• Diskussionen auf der Meta/-Beziehungsebene

• Phasen ohne bzw. mit weniger Kommunikation

3.2.2 Organisatorische Aspekte
Während der Reflexions-Feedback-Sitzungen wurden Themen diskutiert, die die Paarprogrammierung beeinflussen, allerdings nicht einer individuellen sondern einer organisatorischen Ebene zuzuordnen sind.

• Störung durch Dritte (direktes Ansprechen)
Drei Paare wurden während der PP-Sitzung durch Dritte gestört, die mit Fragen oder Probleme auf einen der beiden Programmierer zukamen. Die Programmierer widmeten sich dann der Beantwortung der Fragen. Dies führte allerdings bei zwei Paaren dazu, dass sie in der Konzentration und in ihrem Flow unterbrochen wurden.

• Störung durch Dritte (Lärm)

• Störung aufgrund technischer Probleme
Technische Probleme treten im Bereich der Programmierung immer mal wieder auf, allerdings beeinflussen diese Probleme die Konzentration und den Flow der Paare. Technische Probleme traten in Form von Eclipse oder auch SVN-Problemen auf.

• Konflikt zwischen in individuellen und Teamaufgaben
Einzelne Programmierer wünschen PP gerne häufiger einsetzen, allerdings haben alle Entwickler auch noch ihre individuellen Aufgabenbereiche. Dort entsteht der Konflikt
zwischen der Bearbeitung der individuellen und der Teamaufgaben. Da der Erfolg/Misserfolg der Bearbeitung der individuellen Aufgaben direkt auf den einzelnen Mitarbeiter zurückgeführt werden kann, werden die individuellen Aufgaben bevorzugt und somit weniger PP durchgeführt.

- **Synchronisation der Entwickler**

- **Coding/Working Standards**

  Bemerkung: Im Rahmen der PP-Workshop-Woche ist dieses Problem innerhalb einer Session nicht aufgetreten (ein solches Paar hat nicht gemeinsam programmiert), der Aspekt wurde aber dennoch im Rahmen einer Feedback-Sitzung angesprochen.

### 3.3 Erkenntnisse aus der Abschluss-Runde

#### 3.3.1 Nutzen des Workshops

In der Abschlussrunde wurde der Nutzen des Workshops diskutiert. Für alle Entwickler bei war die Reflexion eine sehr positive und interessante Erfahrung, die ihnen ermöglichte sich über ihre positiven wie auch zum Teil problematischen Verhaltensweisen in der PP bewusst zu werden und diese im Anschluss zu verändern und somit effektiver zusammenzuarbeiten. Speziell das externe Feedback von der FUB auf den PP-Prozess war für die Entwickler mit neuen Einsichten verbunden.

Die Entwickler wurden sich dabei der folgenden Aspekte bewusst:

- Welche Probleme treten in meinem PP-Prozess auf?
  - Wie kann ich diese beheben?
- Welche positiven Verhaltensweisen treten in meinem Prozess auf?
- Wie nimmt mein Partner das Programmieren mit mir wahr?
- Was schätze ich an meinem Partner?
- Warum setzte ich PP ein? Was bringt mir das?
- Welche äußeren Faktoren beeinflussen meinen PP-Prozess?

Da in der Reflexions- und Feedback-Runde individuelle Themen der einzelnen Paare besprochen wurden, konnten die Entwickler unterschiedliche Erkenntnisse aus der Runde mitnehmen. Allgemein können aber die folgenden Richtlinien (ohne organisatorische Erkenntnisse) für den PP-Prozess abgeleitet werden:

- Kommuniziere mit dem Partner
- Diskutiere Entscheidungen
- Teile Dein Wissen
- Greif Dir nicht die ganze Zeit die Maus und Tastatur – der Partner möchte auch in Ruhe tippen
• Etablieren einen aktiven Wechselprozess (für weniger eingespielte Paare)
• Mach Pausen
• Sprich Probleme offen an


Bislang konnte der langfristige Nutzen des Workshops noch nicht evaluiert werden. Hierfür wird noch ein Fragebogen an die Entwickler ausgegeben. Zusätzlich wäre es sinnvoll einen solchen Workshop zu wiederholen, um die langfristigen Veränderungen zu untersuchen.

3.3.2 Feedback zur Organisation des Workshops
Das Feedback zur Organisation war sehr positiv. Speziell, dass die Aufzeichnung für die einzelnen Programmierer wenig zusätzlichen Aufwand bedeuteten und dass die technische Infrastruktur gut funktionierte. Es gab allerdings auch Anregung zur Verbesserung:
• Längere Aufzeichnungsduauer
Die Länge der PP-Sitzungen war auf 1,5 bis 2 Stunden begrenzt. Einige Programmierer setzte diese Einschränkung unter Zeitdruck. Daher sollte die Aufzeichnungsduauer flexibler gestaltet werden können.
• Selbstständiges Starten der Aufzeichnung
Vor jeder Aufzeichnung mussten die Programmierer den Start mit der FUB absprechen. Von Seiten der Programmer wäre es allerdings wünschenswert, die Aufzeichnung ohne Absprache starten zu können.
• Einsatz des Stativs für die Webcam
Die Webcam war auf einem Stativ befestigt, dass auf dem Schreibtisch des Aufzeichnungsrechners stand. Es wäre sinnvoll ein größeres Stativ auf den Boden zu stellen.
• Fokus der Webcam
Positiv wurde angemerkt, dass das Sichtfeld der Webcam hauptsächlich die Tastatur und die Maus und dadurch die Hände der Entwickler aufzeichnete nicht schwerpunktmäßig die Gesichter. Dadurch fühlten sich einige Programmierer weniger beobachtet. Das bewusste Wahrnehmen der Webcam varierte zwischen den Entwicklern jedoch sehr stark. Einige fühlten sich sehr beobachtet (das kann auch das Verhalten beeinflussen) andere nahmen die Webcam nach den ersten 5 Minuten nicht mehr wahr.
• Größerer Schreibtisch
Für die PP sollte ein großer Schreibtisch zur Verfügung gestellt werden, so dass beide Entwickler ausreichend Platz haben.
• Sehr begrenzte Anzahl von Aufzeichnungsmöglichkeiten

4 Zusammenfassung und Ausblick
Im Zuge des Workshops wurden sechs Paare aufgezeichnet und analysiert. Die Entwickler reflektierten über ihre Verhaltensmuster beim PP. Für die Entwickler war der Workshop eine sehr interessante und positive Erfahrung.

Laura Plonka - Freie Universität Berlin
Bei der Reflexion und Analyse wurden positive, wie auch negative Verhaltensmuster im PP sowie externe Einflussfaktoren identifiziert und diskutiert. Speziell bei der Diskussion der paarspezifischen Aspekte konnten die Entwickler Aufschluss über ihr eigenes Verhalten bekommen sowie Anregungen zu mögliche Verbesserungen ihres PP-Prozesses durch Hinweise auf alternative Verhaltensmuster bekommen. Die Identifikation der organisatorischen Herausforderungen wird das Team unterstützen, das Umfeld für die Paar Programmierer weiter zu verbessern.

Derzeit erfolgt eine Detailanalyse der Videos, in der die oben genannten Aspekte genauer im Videomaterial nachvollzogen werden. Sobald aus dieser Analyse Ergebnisse vorliegen, werden diese an weitergegeben.


Sehr interessant wäre auch die Durchführung eines weiteren Workshops, um unterschiedliche Paarkonstellationen zu untersuchen und die Auswirkungen des ersten Workshop zu evaluieren.

5 Referenzen