

University of Groningen

Human-computer interaction in radiology

Jorritsma, Wiard

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Jorritsma, W. (2016). *Human-computer interaction in radiology*. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

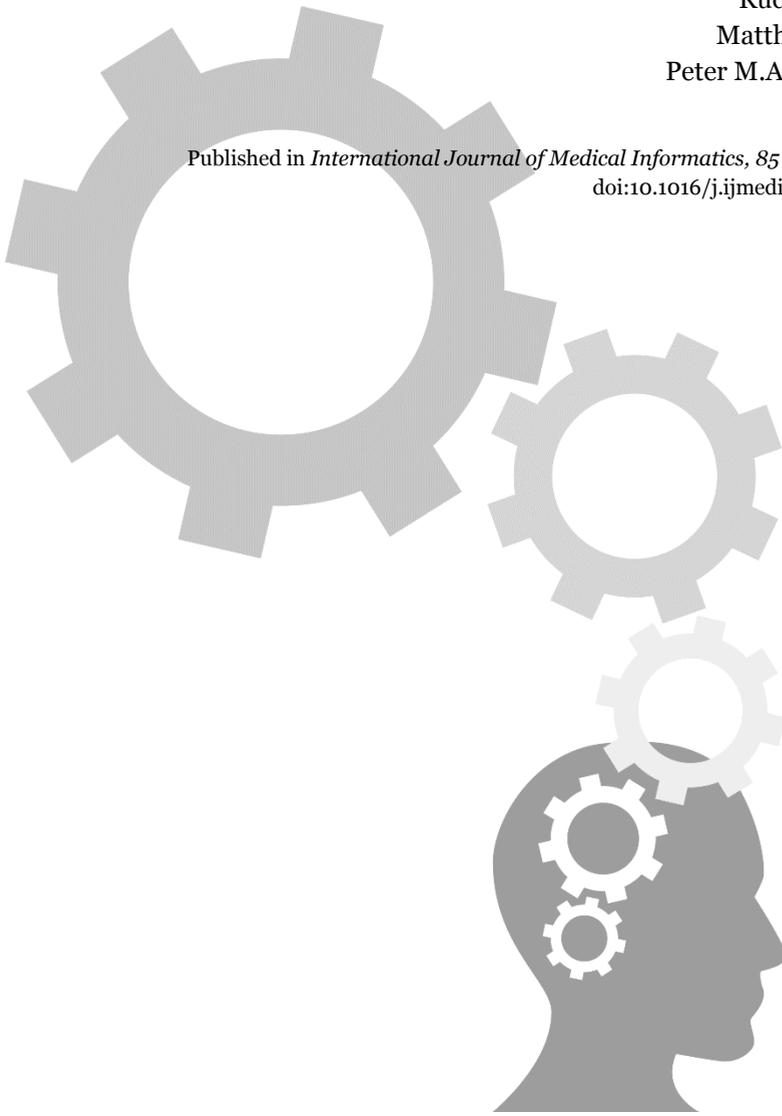
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client

Wiard Jorritsma
Fokie Cnossen
Rudi A. Dierckx
Matthijs Oudkerk
Peter M.A. van Ooijen

Published in *International Journal of Medical Informatics*, 85 (2016), 26–42.
doi:10.1016/j.ijmedinf.2015.10.007



Abstract

Objective: To perform a post-deployment usability evaluation of a radiology Picture Archiving and Communication System (PACS) client based on pattern mining of user interaction log data, and to assess the usefulness of this approach compared to a field study.

Methods: All user actions performed on the PACS client were logged for four months. A data mining technique called closed sequential pattern mining was used to automatically extract frequently occurring interaction patterns from the log data. These patterns were used to identify usability issues with the PACS. The results of this evaluation were compared to the results of a field study based usability evaluation of the same PACS client.

Results: The interaction patterns revealed four usability issues: (1) the display protocols do not function properly, (2) the line measurement tool stays active until another tool is selected, rather than being deactivated after one use, (3) the PACS's built-in 3D functionality does not allow users to effectively perform certain 3D-related tasks, (4) users underuse the PACS's customization possibilities. All usability issues identified based on the log data were also found in the field study, which identified 48 issues in total.

Conclusion: Post-deployment usability evaluation based on pattern mining of user interaction log data provides useful insights into the way users interact with the radiology PACS client. However, it reveals few usability issues compared to a field study and should therefore not be used as the sole method of usability evaluation.

“Data! Data! Data!” he cried impatiently. “I can’t make bricks without clay!”

– Arthur Conan Doyle, The Adventures of Sherlock Holmes

Introduction

Post-deployment usability evaluation, in which a product's usability is evaluated during routine use in a real-world environment, is a valuable addition to pre-deployment usability engineering efforts [1–4]. It is especially useful in domains such as radiology, where it is difficult to perform valid pre-deployment usability evaluations due to difficulties in obtaining a representative user group for usability tests and constructing a representative testing environment [5]. In a recent post-deployment usability evaluation of a radiology workstation [5], we found that radiologists encountered a large number and wide variety of usability issues when using a commercially available radiology workstation in clinical practice. Many of these issues could not realistically have been identified before deployment of the workstation.

This usability evaluation was based on interviews and observations of users during their daily work. While this is a valid and widely used method for post-deployment usability evaluation, it has two important limitations. First, it is difficult to collect large amounts of data from a large sample of users. Second, it is time-consuming to analyze the data that this method yields (e.g. spoken interviews, video recordings of user interaction).

A solution for these problems may lie in a different usability evaluation method: (semi-) automatic analysis of user interaction log data. The basic idea behind this method is to log the actions that users perform on the system and extract usability-related information from these log data. The main advantage of this approach is that it is easy to obtain enormous amounts of data from a large and geographically separated group of users. However, this strength is also its weakness: how to extract meaningful information from these voluminous and low-level log data?

Several solutions to this problem have been proposed (see [6] for an excellent overview). For example, generating counts and summary statistics [7,8], automatically extracting interaction patterns from the log data [9,10], and visualizing the log data in a way that facilitates human interpretation [11,12]. However, there is little empirical research on the relative effectiveness of these methods for usability evaluation. Several studies (e.g. [13–17]) have shown that usability issues could be identified based on user interaction log data, but it is unclear how effective this approach is compared to other methods of usability evaluation.

In this study, we performed a post-deployment usability evaluation of a client for the Picture Archiving and Communication System (PACS), which is the main software component of the radiology workstation, based on an analysis of user interaction log data. We used a data mining technique called *closed sequential pattern mining* [18,19] to automatically extract frequently occurring interaction

patterns from the log data. These patterns were used to identify usability issues with the PACS. In order to assess the relative usefulness of this approach, we compared the results of this evaluation to the results of a field study based usability evaluation of a radiology workstation that included the same PACS client [5].

Methods

Apparatus

Fig. 1 shows the graphical user interface of the PACS client evaluated in this study. At the top of the screen is a toolbar that consists of multiple tabs, each containing a different set of functions. Users can add frequently used functions to the customizable tab labeled *My Tab*, or to the upper region of the right click pop-up menu.

Below the toolbar are one or more *viewports*, which contain the medical images. When the user opens a patient *study* (a collection of one or more series of images obtained during one scanning procedure), a *display protocol* determines the initial viewport layout and arrangement of images based on the study characteristics (e.g. imaging modality, body part scanned) and the user's preferences. Users can manually arrange the images using the *Patient Mini Archive*

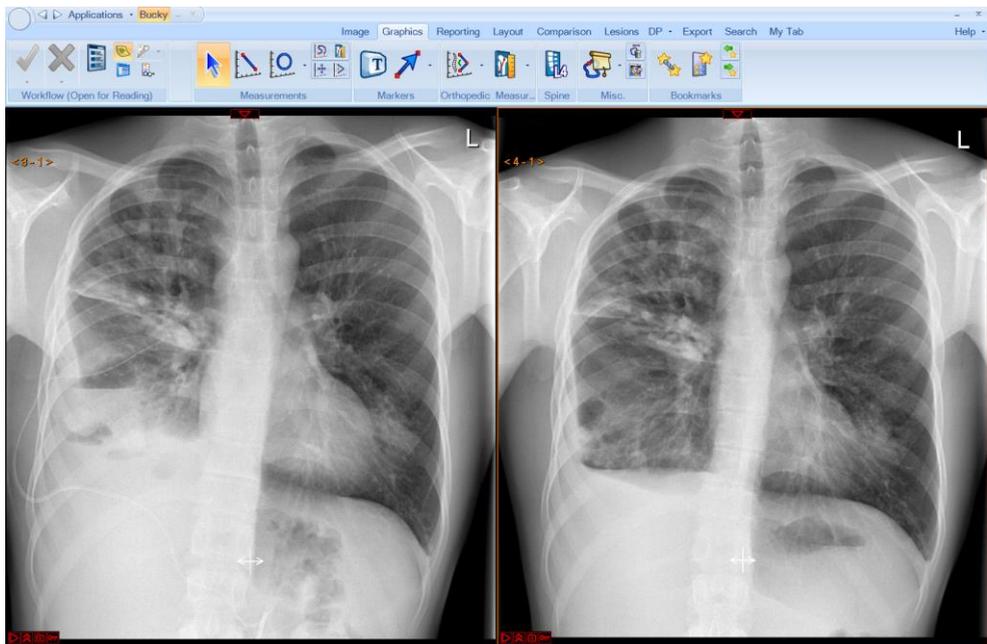


Figure 1. The graphical user interface of the PACS client evaluated in this study.

(PMA), which can be opened in a separate window or displayed as a strip at the bottom of the screen.

Data logging procedure

A modified version of the PACS's built-in logging tool was used to log actions users performed on the PACS client. The following actions were logged: the start and end of a study review; the study description, modality and body part of the study reviewed; the matching display protocol; function selections and their mode of selection (normal toolbar, customizable toolbar, right click menu, or shortcut key); toolbar tab selections; mouse clicks and their coordinates; and the username of the user logged in on the workstation.

All radiologists and radiology residents working in our hospital (a 1339-bed university hospital) were informed of the data logging procedure and were given an opt-out option, which meant that their data would be deleted. This opt-out procedure had been approved by our hospital's legal department before the start of the study. Three out of 69 users chose to opt-out.

Logging started when the PACS client had been in operation in the hospital for eleven months. All users had received equal PACS training.

A script was used to deploy the logging tool on the PACS workstations in our hospital. The logging tool could not be deployed if the workstation was switched off or if the PACS client was running. The script was run at three different time points to maximize the number of successful deployments. Deployment succeeded on 37 out of 64 workstations.

The log files were stored locally and were retrieved from the workstations after four months of logging. A script was used to combine the log data into a single database. Data of opt-out users were removed. Patient information (Patient ID, Accession number, Study UID, Series number and Image UID) was also removed. Usernames were replaced with numbers.

Data analysis

From the log database, a database of *interaction sequences* was generated. An interaction sequence spanned one study review and consisted of the type of study reviewed (modality + body part) followed by all actions the user performed during the review. The left/right modifiers of the body part were removed, which meant that for example x-rays of the left and right hand were considered the same type of study. Display protocol information and mouse clicks were not included in the sequences, because they were not considered meaningful in preliminary analyses. In the vast majority of cases, the display protocol information was redundant with the study description, modality and body part of the study reviewed. Mouse clicks on interface buttons were redundant with function selections. Mouse clicks on

images were meaningless because they did not specify which images was clicked on. This could also not be inferred from the coordinates of the click, because we did not know how the images were arranged on the screen. The resulting sequence database contained 10,948 interaction sequences of 66 users.

A data mining technique called *closed sequential pattern mining* was used to discover frequently occurring interaction patterns in the log data. Given a sequence database, this technique generates all *closed sequential patterns* above a predefined *support* threshold. A *sequential pattern* is a sequence of actions (not necessarily consecutive) and the *support* of a pattern is the number of interaction sequences in which the pattern occurs. A sequential pattern is *closed* if it has no super-pattern (a longer pattern that includes it) with the same support.

We illustrate this with an example. Given the following sequence database:

MR skull → Line measurement → Mark as key image

CT thorax → Window level tool → Oval ROI → Mark as key image

CT thorax → MPR → Sagittal plane → Line measurement → Mark as key image

closed sequential pattern mining with a minimum support of 2 yields the following patterns:

Pattern	Support
Mark as key image	3
Line measurement → Mark as key image	2
CT thorax → Mark as key image	2

Note that patterns can consist of only one action and that actions do not have to be consecutive in order to form a pattern. The pattern Mark as key image has a support of 3 because it occurs in 3 interaction sequences. The patterns Line measurement and CT thorax are above the support threshold (their support is 2), but they are not included because they have super-patterns with the same support (i.e. they are non-closed).

The closed sequential pattern mining approach is similar to using other data mining techniques (e.g. association rule mining [20,21], sequential pattern mining [22,23], maximal repeating pattern analysis [9]), but it has some advantages. First, it only mines closed patterns, which leads to clean output (without irrelevant non-closed patterns). Second, the user actions within the patterns it generates are

sequential, but not necessarily consecutive. This allows for the detection of patterns in which an action *eventually* leads to another action. It does not matter how many irrelevant actions occur in between.

We used the CM-ClaSP algorithm of the open-source SPMF data mining library [20] to perform the closed sequential pattern mining. The ClaSP algorithm has been shown to outperform other algorithms for this task [19] and CM-ClaSP is a modified version of ClaSP that is even more efficient [21].

We ran the algorithm with a minimum support of 200 (i.e. the patterns should occur in at least 200 study reviews), which produced 62 closed patterns. We marked the patterns that indicated suboptimal user interaction, which was defined as a pattern or a set of patterns that violated our expectations about effective or efficient use of the PACS client based on our knowledge of the PACS client's user interface.

To acquire more information about the suboptimal user interaction, we ran the algorithm again with a support of 50, which produced 633 closed patterns. For all marked patterns, we searched these 633 patterns for related patterns (super-patterns or patterns containing related user actions) that could provide additional information on the interaction. We used our knowledge of the PACS client's user interface to infer usability issues from the interaction patterns.

The results of this evaluation were compared to the results of a field study based usability evaluation of a radiology workstation that included the same PACS client [5]. This evaluation consisted of interviews and observations of users during their daily work and identified 92 usability issues with the radiology workstation. 48 of these issues occurred in the PACS client.

During the identification of usability issues based on the interaction patterns, we did not have access to the results of the field study.

Results

Table 1 shows the most frequently occurring closed interaction patterns (with a minimum support of 200). It is important to recall that the sequence database consisted of 10,948 study reviews and that the support of a pattern represents the number of reviews in which this pattern occurred. The marked patterns indicate suboptimal user interaction. The inference of usability issues from the interaction patterns is described in the following paragraphs.

Patterns 7, 23 and 50 show that users often open the patient mini archive (PMA), which is used to manually arrange the patient images on the screen. This indicates that the display protocols, which automatically arrange the images on the screen based on their characteristics and the user's preferences, do not function properly.

Table 1. The most frequently occurring closed interaction patterns (with a minimum support of 200). The marked patterns indicate suboptimal user interaction. Patterns marked with the same color belong to the same interaction.

	Pattern	Support
1	DX thorax	1193
2	CT thorax	996
3	Line measurement	982
4	CT abdomen	960
5	MPR	926
6	US abdomen	807
7	Open PMA	784
8	CT skull	778
9	Window level preset 4	673
10	Pointer	604
11	Coronal plane	542
12	Window level preset 8	535
13	DX hand	519
14	Sagittal plane	505
15	Zoom from cursor	500
16	MPR → Coronal plane	452
17	Pan	444
18	Window level preset 7	440
19	MPR → Sagittal plane	425
20	MR Skull	416
21	Mark as key image	386
22	Window level preset 3	377
23	Open PMA → Open PMA	371
24	Axial plane	360
25	Open external 3D application	356
26	Graphics tab	326
27	Layout tab	314
28	Export tab	313
29	Line measurement → Pointer	304
30	Line measurement → Line measurement	280
31	MPR → Axial plane	279
32	MPR → MPR	273
33	Window level preset 4 → Window level preset 4	263
34	CT total body	258
35	Window level preset 1	256
36	Export tab → Open external 3D application	253
37	Window level tool	244
38	CT thorax → Window level preset 4	240
39	Window level preset 4 → Window level preset 3	238
40	MR thorax	231
41	Viewer rendition	229
42	Window level preset 4 → Window level preset 7	229
43	Coronal plane → Axial plane	229
44	Mark as key image → Mark as key image	226
45	MR abdomen	223

46	MPR → Window level preset 8	222
47	CT abdomen → MPR	221
48	DX pelvis	221
49	CT spine	215
50	Open PMA → Open PMA → Open PMA	215
51	Window level preset 7 → Window level preset 7	212
52	MIP	211
53	Window level preset 8 → Window level preset 8	210
54	Window level preset 4 → Window level preset 8	210
55	Automatic registration	210
56	Window level preset 9	207
57	Coronal plane → Coronal plane	206
58	DX spine	205
59	DX foot	202
60	Window level preset 6	202
61	Line measurement → Mark as key image	200
62	Coronal plane → Sagittal plane	200

Note that the logging tool only logged the action of opening the PMA. The actual PMA usage frequency was therefore higher than the pattern data suggest, because many users use the PMA in strip mode or display it on a second monitor which means they do not have to open it every time they need it.

Pattern 10 shows that users often selected the pointer function, which changes the function of the left mouse button back to the default mouse cursor. The reason for this is that functions that change the mouse cursor behavior (e.g. line measurement, zoom, pan) stay active until another function that changes the mouse cursor behavior is selected. Users select the pointer function, which in itself is useless, to deactivate other functions.

Pattern 29 shows that users often use the pointer to deactivate the line measurement function. The fact that the line measurement function stays active until it is deactivated could be useful when users have to perform multiple consecutive measurements. However, the pattern data show that performing one line measurement in a review is more common than performing multiple line measurements (Line measurement, support = 982; Line measurement → Line measurement, support = 280; Line measurement → Line measurement → Line measurement, support = 110). In the 982 reviews in which users made a line measurement, they only made more than one line measurement in 280 reviews. Therefore it would be better to automatically deactivate the function after one use.

Pattern 25 shows that users often open an external 3D application. Most of this application's functionality is also available within the PACS. The fact that users spend additional time and effort to load the patient images into this application

indicates that the PACS's built-in 3D functionality does not allow users to effectively perform certain 3D-related tasks.

Patterns 26, 27 and 28 show that users often switched to other tabs in the toolbar. This indicates that they underuse the PACS's possibility to add frequently used functions to the customizable toolbar tab (My Tab) and right click menu, which would eliminate the need to switch between the tabs. Pattern 36 shows an example of a function that would be useful to have in the customizable toolbar for many users. Of the 356 times users opened the external 3D application, they did so by first switching to the export tab, in which this function is located, 253 times.

Table 2 shows the usability issues identified based on the log data analysis and related issues identified in the field study [5]. All usability issues identified based on the log data were also found in the field study. The field study revealed more detailed information about the issues regarding the display protocols and 3D functionality. In total, four usability issues were identified based on the log data analysis compared to 48 in the field study.

Table 2. Usability issues identified based on the log data analysis and related issues identified in the field study.

Log data analysis – usability issues	Field study – related usability issues
<ul style="list-style-type: none"> The display protocols do not function properly. 	<ul style="list-style-type: none"> There are no appropriate display protocols for some types of studies (e.g. brain MRI and musculoskeletal studies). Display protocols treat digital radiography and computer radiography studies as different studies, while they are equivalent to the user. Display protocols do not recognize relevant images within a study that consists of multiple body parts (e.g. CT abdomen images within a CT total body study). Display protocols do not recognize studies received from other hospitals. If there is a study (e.g. a lumpectomy) between the current and the previous mammogram, the mammography display protocol displays this study instead of the previous mammogram.
<ul style="list-style-type: none"> The line measurement tool stays active until another tool is selected, while it would be better to deactivate it after one use. 	<ul style="list-style-type: none"> Measurement tools stay active until another tool is selected. This requires an additional user action and often leads to accidental measurements.

- The 3D functionality does not allow users to effectively perform certain 3D-related tasks.
 - The scroll bar disappears when switching between 2D and multi-planar reconstruction (MPR)¹ mode.
 - When two series have been registered and at least one of them is an MPR, scrolling in the MPR series results in a lag in the other series.
 - Getting two multi-view MPRs next to each other is cumbersome and counterintuitive.
 - In multi-view MPR mode, when users adjust the angulation of the planes and drag a different plane into the main viewport, the angulation is reset.
 - In multi-view MPR mode, users can only rotate the plane that is in the main viewport. This means that they have to drag a plane into the main viewport in order to rotate it.
 - The planes in multi-view MPR mode move independently of each other. This makes it difficult for users to judge whether they have adjusted the angulation correctly.
 - Switching between planes in multi-view MPR mode is inconvenient. Users have to drag the plane they want to view to the main viewport by a small action region in the top left of the plane.
 - In multi-view MPR mode, the plane rotation symbol often fails to appear, which means that users cannot rotate the plane. If users move the plane once, the rotation symbol does appear.
 - There is a significant lag when interacting with 3D reconstructions, especially when zoomed in.
-
- Users underuse the customization possibilities of the My Tab/right click menu.
 - Many users are unaware of the My Tab's customization possibilities, while they indicated that they would find them useful.

Discussion

In this study, we performed a post-deployment usability evaluation of a PACS client based on an analysis of user interaction log data. We used a data mining technique called closed sequential pattern mining to automatically extract frequently

¹ MPR is the most frequently used 3D functionality. The PACS evaluated in this study has two MPR modes, which we refer to as MPR and multi-view MPR. The former displays one anatomical plane at a time and allows users to switch between the planes. The latter displays the three anatomical planes simultaneously.

occurring interaction patterns from the log data. These patterns were used to identify usability issues with the PACS. In order to assess the relative usefulness of this approach, we compared the results of this evaluation to the results of a field study based usability evaluation of the same PACS client [5].

The log data analysis revealed four usability issues with the PACS, compared to 48 issues that were identified in the field study. Previous usability evaluations based on log data analysis also identified modest numbers of usability issues (fourteen issues with an e-mail application [14], nine with an e-learning application [13], two with the Ryanair and Journey Planner websites [16], and one with the Skyscanner and TU/e websites [16]). However, because these studies used log data analysis as the sole method of usability evaluation, they do not provide insight into the relative effectiveness of this approach. Our results indicate that log data analysis is an ineffective approach for identifying usability issues compared to a field study.

The field study also yielded more detailed information about the usability issues than the log data analysis. Two of the four issues identified based on the log data were actually compound issues consisting of multiple more specific issues. These compound issues are not directly actionable (e.g. what changes need to be made to address the issue that ‘the display protocols do not function properly?’) and therefore require a more in-depth follow-up evaluation.

Considering the limited resources available for usability evaluation, a log data analysis could be useful as a preliminary analysis that guides the collection of other sources of usability data. This could focus other usability evaluation methods on aspects of the system with which a large number of users exhibit suboptimal interaction patterns.

For the issue regarding the continued activation of the measurement tools, the log data analysis was more actionable than the field study. Comments from the PACS vendor revealed that the activation behavior of the measurement tools could be customized, thereby solving the issue (none of the users in the field study were aware of this possibility). The data from the field study suggested that it might be more appropriate to change the default setting of the measurement tools to ‘deactivate after one use’. However, because this has negative consequences when users perform multiple consecutive measurements, and the field study did not yield any information about whether single or consecutive measurements were more common, the optimal setting could not be determined. The log data analysis on the other hand, did show that single measurements are far more common than multiple measurements, and suggests that the optimal setting is therefore ‘deactivate after one use’.

An interesting possibility that could solve usability issues caused by suboptimal configuration is to use the log data to automatically provide user-specific

customization support. For example, based on users' log data, the PACS could automatically suggest that users should add a specific frequently used function to the customizable toolbar, or that they should change the default activation behavior of the measurement tools. If user interaction related to the arrangement of images on the screen would be logged, the PACS could even suggest the most effective display protocol for a given study for a specific user. A preliminary version of such adaptive customization support in a radiology setting already produced promising results [22]. However, more research is needed to truly determine the effectiveness of this approach.

An important limitation of log data analysis is that it reveals what users are doing, but not why they are doing it. The log data therefore leave a lot of room for interpretation. For example, we interpreted the high usage frequency of the PMA as an indicator that the display protocols were not functioning properly. This interpretation was confirmed by the field study, but it could very well have been incorrect. Maybe users just prefer to arrange the images manually. Or maybe they open the PMA because it contains important information about the patient (although this does not explain why they open it multiple times during one study review). We believe that this difficulty in inferring the cause of interaction patterns makes it vital to supplement log data analysis with observational usability evaluation methods, in order to check whether the conclusions drawn from the log data are actually valid.

Another limitation is that the scope and nature of usability issues that can be identified based on log data analysis depend on the quality and level of abstraction with which the interaction information was logged [6]. Some aspects of interaction are difficult, if not impossible, to log properly (e.g. interaction with images), which seriously hampers the ability of log data analysis to detect issues with these interaction aspects.

To the best of our knowledge, this study was the first to use the closed sequential pattern mining technique for usability evaluation. This approach is similar to using other data mining techniques (e.g. association rule mining [23,24], sequential pattern mining [25,26], maximal repeating pattern analysis [9]), but it has some advantages. First, it only mines closed patterns, which leads to clean output (without irrelevant non-closed patterns). Second, the user actions within the patterns it generates are sequential, but not necessarily consecutive. This allows for the detection of patterns in which an action *eventually* leads to another action. It does not matter how many irrelevant actions occur in between. We found these properties useful during our usability evaluation.

A difference between our study and previous usability evaluations based on log data analysis is that they analyzed log data of users performing pre-defined tasks (nine tasks in [14], six in [13] and one to three in [16]), whereas we analyzed the

data of users performing their daily work in an unconstrained fashion. The advantage of using pre-defined tasks is that the tasks corresponding to users' interaction patterns are known. This simplifies the data interpretation, because the observed interaction patterns for a given task can be compared to the optimal pattern to perform this task. Deviations from the optimal patterns then indicate usability issues (note that these still have to be verified with observational data). However, the disadvantages are that the evaluation is based on artificial rather than naturally occurring tasks, users have to take time off from their daily work to perform the tasks, and the evaluation cannot identify issues with aspects of the system that were not interacted with during the tasks. We believe that these disadvantages outweigh the advantage and that logging natural interaction is therefore preferable over logging task-based interaction. We have demonstrated that meaningful interaction patterns can be detected in log data of natural interaction.

A potential source of bias in our study is that we performed the pattern mining on a single database consisting of the interaction sequences of all users. This means that the mined patterns could theoretically come from a single user that was therefore overrepresented in the interaction sequence database.

In the design phase of the study, we considered an alternative approach to eliminate this potential bias: (1) construct a database of interaction sequences for each individual user; (2) for each user, perform the pattern mining on his or her database with a minimum support of x % of the total number of interaction sequences in the database, where x is an empirically determined constant that is the same for all users; (3) combine the patterns of the individual users into a single set of patterns.

While this approach eliminates the possibility that all patterns come from a single user (or a handful of users), it introduces several new problems. First, it is possible that meaningful patterns are missed in step 2. For example, consider a pattern that occurs for all users, but falls just below the support threshold for each individual user. This pattern would not be included in the combined set of patterns in step 3, whereas it would have been included if the pattern mining was done on a combined sequence database of all users. Second, the combination of the patterns of the individual users into a single set of patterns in step 3 also introduces a potential bias. If equal weight is given to the patterns of each user, users with a small sequence database (who maybe only used the PACS client once) are overrepresented. On the other hand, if the weight given to each user's patterns is proportional to the size of his or her sequence database, we have a similar bias to the one we were trying to eliminate, namely that users with a large number of interaction sequences are overrepresented in the combined set of patterns. This approach also has the practical disadvantage that it makes the pattern mining

process more complex and more computationally expensive. We concluded that the new problems this approach introduces do not outweigh the potential bias it prevents. We therefore decided against this approach.

A limitation of our study is that we only evaluated the usability of one application, we only used one log data analysis method, and we only compared the log data analysis to one other usability evaluation method (a field study). Although this provided meaningful insights into the relative usefulness of the log data analysis approach, more empirical research on log data analysis for usability evaluation is needed to obtain a more robust body of results. Future research should ideally compare multiple methods of log data analysis to multiple other usability evaluation methods for a variety of applications. Although it is not feasible to cover all of these aspects in one study, each study that covers some of them adds to the amount of empirical data on the usefulness of log data analysis for usability evaluation.

Although the usability issues found in this study are most relevant for the particular application we evaluated, they are definitely not limited to this application alone. The data on the display protocol issue are relevant for any PACS that uses display protocols, the data on the line measurement issue are relevant for any application that allows users to measure on images, and the finding that users underuse the customization facilities is relevant for any application with an adaptable interface. Furthermore, our study can serve as a useful guide to people who want to use log data analysis for usability evaluation of any application.

Conclusion

Post-deployment usability evaluation based on pattern mining of user interaction log data provides useful insights into the way users interact with the radiology PACS client. However, it reveals few usability issues compared to a field study and should therefore not be used as the sole method of usability evaluation. It can however be a useful addition to other usability evaluation methods, because the log data can be used to guide the collection of other sources of usability data, to obtain an accurate estimate of an issue's frequency of occurrence, and to determine the optimal default configuration settings.

References

- [1] J. Nielsen, The usability engineering life cycle, *Computer*. 25 (1992) 12–22.
- [2] P.K. Chilana, A.J. Ko, J.O. Wobbrock, T. Grossman, G. Fitzmaurice, Post-deployment usability: a survey of current practices, in: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, ACM Press, Montreal, Canada, 2011: pp. 2243–2246.

- [3] D.M. Nichols, D. McKay, M.B. Twidale, Participatory Usability: supporting proactive users, in: Proc. 4th Annu. Conf. ACM Spec. Interes. Gr. Comput. Interact., Dunedin, New Zealand, 2003: pp. 63–68.
- [4] C.M. Nielsen, M. Overgaard, M.B. Pedersen, J. Stage, S. Stenild, It's worth the hassle! The added value of evaluating the usability of mobile systems in the field, in: Proc. 4th Nord. Conf. Human-Computer Interact., ACM Press, Oslo, Norway, 2006: pp. 272–280.
- [5] W. Jorritsma, F. Cnossen, R. Dierckx, M. Oudkerk, P.M.A. van Ooijen, Post-deployment usability evaluation of a radiology workstation, *Int. J. Med. Inform.* 85 (2016) 28–35.
- [6] D.M. Hilbert, D.F. Redmiles, Extracting usability information from user interface events, *ACM Comput. Surv.* 32 (2000) 384–421.
- [7] D.R. Olsen, B.W. Halversen, Interface usage measurements in a user interface management system, in: Proc. 1st Annu. ACM SIGGRAPH Symp. User Interface Softw., ACM Press, Banff, Canada, 1988: pp. 102–108.
- [8] E. Chang, T.S. Dillon, Automated usability testing, in: S. Howard, J. Hammond, G. Lindgaard (Eds.), Proc. IFIP TC13 Interantional Conf. Human-Computer Interact., Chapman and Hall, Sydney, Australia, 1997: pp. 77–84.
- [9] A.C. Siochi, D. Hix, A study of computer-supported user interface evaluation using maximal repeating pattern analysis, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst., ACM Press, New Orleans, LA, 1991: pp. 301–305.
- [10] A. Vargas, H. Weffers, H.V. da Rocha, A method for remote and semi-automatic usability evaluation of web-based applications through users behavior analysis, in: Proc. 7th Int. Conf. Methods Tech. Behav. Res., ACM Press, Eindhoven, The Netherlands, 2010: pp. 19:1–19:5.
- [11] M. Gray, A. Badre, M. Guzdial, Visualizing usability log data, in: Proc. 1996 IEEE Symp. Inf. Vis., IEEE, San Fransisco, CA, 1996: pp. 93–98.
- [12] J.I. Hong, J.A. Landay, WebQuilt: a framework for capturing and visualizing the web experience, in: Proc. 10th Int. Conf. World Wide Web, ACM Press, Hong Kong, Hong Kong, 2001: pp. 717–724.
- [13] A. Vargas, H. Weffers, H.V. da Rocha, Discovering and analyzing patterns of usage to detect usability problems in web applications, in: Proc. 11th Int. Conf. Intell. Syst. Des. Appl., IEEE, Cordoba, Spain, 2011: pp. 575–580.
- [14] H. Okada, T. Asahi, GUITESTER: a log-based usability testing tool for graphical user interfaces, *IEICE Trans. Inf. Syst.* E82-D (1999) 1030–1041.
- [15] V.F. de Santana, M.C.C. Baranauskas, WELFIT: A remote evaluation tool for identifying Web usage patterns through client-side logging, *Int. J. Hum. Comput. Stud.* 76 (2015) 40–49.
- [16] A. Vargas, H. Weffers, H. V Da Rocha, Analyzing user interaction logs to evaluate the usability of web applications, in: 3rd Symp. Web Soc., IEEE, Port Elizabeth, South Africa, 2011: pp. 61–67.
- [17] D. Bader, D. Pagano, Towards automated detection of mobile usability issues, in: Proc. 1st Eur. Work. Mob. Eng., Aachen, Germany, 2013: pp. 341–354.
- [18] X. Yan, J. Han, R. Afshar, CloSpan: Mining closed sequential patterns in large datasets, in: Proc. 3rd SIAM Int. Conf. Data Min., SIAM, San Fransisco, CA, 2003: pp. 166–177.

- [19] A. Gomariz, M. Campos, R. Marín, B. Goethals, ClaSP: an efficient algorithm for mining frequent closed sequences, in: Proc. 17th Pacific Asia Conf. Knowl. Discov. Data Min., Springer-Verlag, Gold Coast, Australia, 2013: pp. 50–61.
- [20] P. Fournier-Viger, A. Gomariz, A. Soltani, H. Lam, T. Gueniche, SPMF: open-source data mining platform. <http://www.philippe-fournier-viger.com/spmf/>. Accessed 21 May, 2014, (2014).
- [21] P. Fournier-Viger, A. Gomariz, M. Campos, R. Thomas, Fast vertical mining of sequential patterns using co-occurrence information, in: Proc. 18th Pacific-Asia Conf. Knowl. Discov. Data Min., Springer, Tainan, Taiwan, 2014: pp. 40–52.
- [22] W. Jorritsma, F. Cnossen, P.M.A. van Ooijen, Adaptive support for user interface customization: a study in radiology, *Int. J. Hum. Comput. Stud.* 77 (2015) 1–9.
- [23] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proc. 1993 ACM SIGMOD Int. Conf. Manag. Data, ACM Press, Washington, D.C., 1993: pp. 207–216.
- [24] P. Fournier-Viger, C.-W. Wu, V.S. Tseng, Mining top-k association rules, in: Proc. 25th Can. Conf. Artif. Intell., Springer-Verlag, Toronto, Canada, 2012: pp. 61–73.
- [25] R. Agrawal, R. Srikant, Mining sequential patterns, in: Proc. Elev. Int. Conf. Data Eng., IEEE, Taipei, Taiwan, 1995: pp. 3–14.
- [26] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, et al., Mining sequential patterns by pattern-growth: the PrefixSpan approach, *IEEE Trans. Knowl. Data Eng.* 16 (2004) 1424–1440.