Towards Standardized Pen-Based Annotation of Breast Cancer Findings

Suzanne Kieffer[1], Annabelle Gouze[1], Ronald Moncarey[1], Christian Van Brussel[1], Jean-François De Wispelaere[2], Françoise Kayser[2], and Benoît Macq[1]

[1] Communications and Remote Sensing Laboratory, Université catholique de Louvain
Place du Levant 2, B-1348 Louvain-La-Neuve, Belgium
[2] Cliniques universitaires UCL de Mont-Godinne
Avenue G.Thérasse 1, B-5530 Yvoir, Belgium
{suzanne.kieffer,annabelle.gouze,ronald.moncarey,
christian.vanbrussel,jean-francois.dewispelaere,
francoise.kayser,benoit.macq}@uclouvain.be

Abstract. The development of computer technologies provides a means to support and facilitate the daily activities of potentially all users. This may be of particular importance for experts in breast cancer imaging and diagnosis. While many research efforts have been carried out separately on the implementation of task-oriented systems, much less effort has been undertaken to design and develop technologies compliant with domain standards or in accordance with end-user needs and expectations. This further suggests the need to improve both the usefulness and the usability of breast cancer-dedicated systems. This paper reports the results of a development method combining the application of user-centered design together with usability development methods. At different time frames in the life-cycle, the development method employed knowledge elicitation interviews, scenario-focused questionnaires, paper mock-ups and usability tests. Owing to its naturalness and its convenience, pen-based interaction with a graphics tablet was chosen as the modality to interact with the system. Additional innovative solutions were designed and implemented in order to facilitate and improve the visualization and the manipulation of data during the lesion characterization: namely an icon framework, a star-menu and a semi-automatic lesion detection system. The resulting user interface is a pen-based interactive tool supporting visualization, navigation, standardized lesion characterization and reporting. The usability tests suggest that it provides end-users with an efficient, reliable and usable system.

1 Context and Motivations

Since digital mammography has replaced screen-film mammography, information technology has progressively been introduced in breast cancer (BC) screening and diagnosis. Owing to the specificity and the complexity of each task involved in mammogram analysis (e.g., image visualization, image analysis, lesion detection, interpretation and reporting), research efforts have been focusing mainly on the implementation of task-oriented systems such as image viewers, computer-aided diagnosis (CAD) software, digital case databases, etc. Hence, radiologists currently

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2009, LNCS 5613, pp. 524–533, 2009.

tend to split their work between a growing number of interactive tools, workstations and media (e.g., screen, mouse, keyboard, and handheld recorder).

In addition, interactive tools for BC screening and reporting have to be compliant with the BI-RADS standard [1], an approved system of descriptive terms and reporting guidelines. Such tools not only facilitate reporting, providing radiologists with structured and standardized reports, but lead to data accessibility as well (e.g., data exchange and storage, interpretation monitoring [13], retrieval of useful and interesting cases for teaching and research purposes [14]).

Finally, interests in user-centered design together with usability stem from the goal to design and implement interactive systems supporting the activities of domain-expert users, who are not necessarily experts in computer science. As highlighted in [3], great care must be devoted to the study of the needs and the expectations of such domain-expert users. In particular, attention must be paid to usability throughout the software life-cycle in order to design and implement user-friendly and easy-to-use interfaces [9,6].

2 Objectives and Significance of the Work

The ultimate objectives of the research presented here are to design, implement and evaluate a BC-oriented interactive system which integrates the interactive annotation of significant findings (i.e., lesion characterization and reporting) and the semi-automatic lesion detection. This paper focuses on the design and the evaluation of the annotation tool; refer to [5] for more details about the semi-automatic lesion detection tool. A prototype for lesion annotation, based on semantic web technologies, was presented during SPIE Medical Imaging 2007 [4]. Our work goes further mainly by providing BC-experts with a new interaction style to characterize findings: the pen-based annotation with a graphics tablet.

The significance of the work can be highlighted depending on three axes:

- End-user: provide experts in BC screening with useful and usable tools;
- Accessibility: increase breast imaging data accessibility thanks to standardization (data exchange and storage, interpretation monitoring [13], useful and interesting cases retrieval for teaching and research purposes [14]);
- Usability: develop and promote support for designers of usable systems.

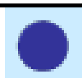
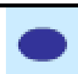
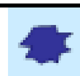
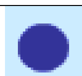
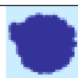
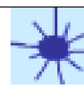
3 Annotation of Breast Cancer Finding

The pen-based user interface (PUI) is an effective method to provide end-users with a natural, intuitive and convenient interaction [12]. Owing to its high naturalness and mainly to its convenience to satisfy the lesion characterization requirement, pen-based interaction with a graphics tablet was chosen as the modality to interact with the system: navigating in a clinical case (i.e., among the mammograms), navigating in a specific mammogram (i.e., zoom-in, zoom-out), sketching a region of interest (ROI), annotating findings and reporting (i.e., direct manipulation of menus, icons, widgets).

The BI-RADS [1] provides a standardized terminology for the description of BC findings. Any finding is described according to a lesion type (i.e., mass, calcification, architectural distortion, special case or associated finding), and type-related characteristics. Beyond the specific characteristics related to a lesion type, the breast imaging report contains the finding location and the comparison to previous studies, whatever the type.

An icon framework was created according to this standard in order to enable any finding to be fully described. Every single term of the BI-RADS lexicon is represented by a unique icon (altogether about 150 different icons), so that the lesion characterization is straightforward and unambiguous. A color code was adopted in order to facilitate the discrimination between the findings: masses in blue, calcifications in yellow, architectural distortions in green, special cases in violet, and associated findings in orange. The schemes on the icons related to the finding location and the comparison to previous studies are common to all the lesion types; only colors are different. The Table 1 presents the icons related to the specific characteristics of masses. Masses are characterized by basic shape (round, oval, lobular or irregular), margin (circumscribed, microlobulated, obscured, indistinct or spiculated) and density (high-density, equal density, low-density or fat-containing radiolucent).

Table 1. Icons related to the specific characteristics of masses: shape, margin and density

Shape	 Round	 Oval	 Lobular	 Irregular	
Margin	 Circumscribed	 Microlobulated	 Obscured	 Indistinct	 Spiculated
Density	 High-density	 Equal Density	 Low-density	 Fat-Containing Radiolucent	

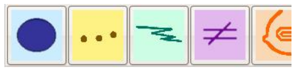
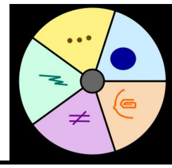
4 Development Method

The development method combined user-centered design together with usability development methods, and employed: knowledge elicitation interviews, scenario-focused questionnaires, paper mock-ups and usability tests.

Domain- and task-relevant knowledge was collected early in the life-cycle thanks to knowledge elicitation interviews. Five domain-expert users were questioned thoroughly about the BC domain, the task series involved in their activity, their needs and their expectations with the goal to implement the collected information in the system. The equipment used was paper notes and video recording.

Scenario-focused development method was used to define and select the interaction scenarios which would best support and fit end-user activities. Thanks to

Table 2. Scenario-focused questionnaire. The end-user activity is the description of the lesion type (column 1). There were two proposed interactive scenarios to support this activity: using an array of buttons or using a pie-menu (respectively, columns 2 and 3).

Screening analysis activity	Scenario#1	Scenario#2
Description of the lesion type: • Masses • Calcifications (CA++) • Architectural distortion • Special cases • Associated findings	Array of buttons: 	Pie menu: 

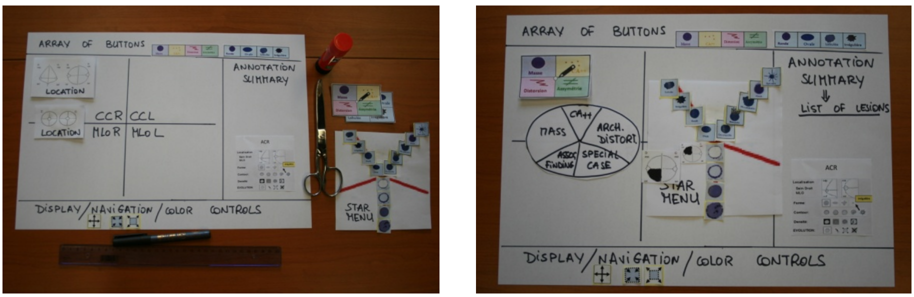


Fig. 1. Paper mock-ups. On the left: the experimental material such as paper, glue, and pen. On the right: the device in use.

end-user involvement, such a design method proved to reduce both development time and development cost, and to improve usability [12]. Referring to [12], a written scenario-focused questionnaire (see Table 2) and paper mockups (see Fig. 1) were implemented complementarily.

Both were elaborated from the information collected during knowledge elicitation interviews and were presented to six domain-expert users in order to evaluate the icon framework and the overall spatial organization of the interface, and to select the potential interactive scenarios. The questionnaire was used by the interviewer as visual aid during face-to-face interviews. Paper mock-ups were preferred to computer prototypes since the available prototyping tool did not support the scenarios applying for implementation, and since the evaluation should lead to a lot of drawings, direct manipulation of paper components and discussions between designers and domain-expert users [10]. The analysis of the data collected from scenario-focused questionnaire and paper mock-ups led to the expert validation of: the terminology and the iconic framework, the color code, and the preliminary expert validation of the Pie and the Star menus.

Providing experts in BC imaging with an interactive tool supporting their activity is a tough problem of human-computer interaction considering the user requirements of usefulness and usability. The usefulness is ensured by the compliance of the system with the domain standard (see section 3), and by the integration within a single interactive tool of the functionalities of image visualization, finding characterization (i.e., annotation), semi-automatic lesion detection and reporting. In order to guarantee the usability of the system, especially during the lesion characterization task, sustained attention has been paid to the graphical representation of the mammographic BI-RADS terminology (by the design of an exhaustive set of icons), to the spatial organization of multimedia data (not only the overall spatial organization of the user interface, but the specific location of widgets as well), and to the design of new interactive solutions suited to the finding annotation with a pen on a graphics tablet. Therefore, two complementary menus were implemented to support the pen-based annotation of BC findings: a pie menu [2] for the pen-based selection of the lesion type, and a star menu for the pen-based description of the type-related characteristics of the lesion.

The pie menu (see Fig. 2, left) was implemented in order to facilitate the pen-based annotation of the lesion type. This format was chosen because it reduces the target seek time and improves the accuracy of target selection [2,8]. The star menu (see Fig. 2, right) was implemented in order to facilitate the pen-based annotation of complementary characteristics by grouping icons related to the same characteristic on a single line. This format was chosen because such a display layout was proved to be very efficient and accurate for visual inspection or visual detection by comparison with matrix, elliptic and random spatial structures [11].

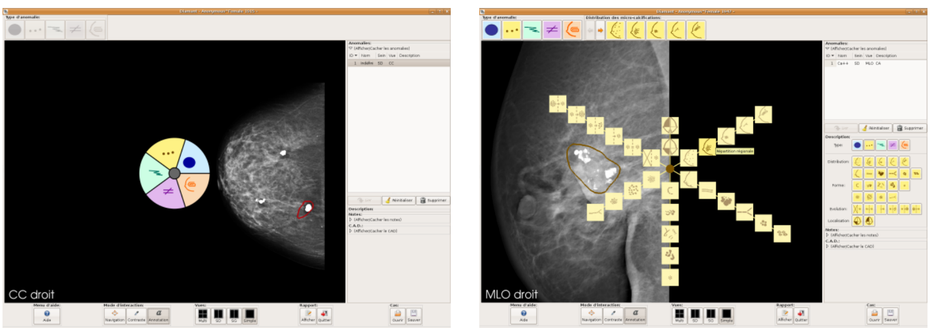


Fig. 2. Menus for the pen-based annotation of findings. Pie menu (left) and star menu (right).

5 Usability Evaluation of the Pie and the Star Menus

The purpose of this investigation was to evaluate the usability of the pie and the star menus (pie-star menus) during the pen-based annotation of BC findings. The usability evaluation criteria were the system effectiveness, the system efficiency and the satisfaction of the users [9,6]. The effectiveness metric was the task completion, whereas the efficiency metrics were task completion times and the number of clicks.

Participants were asked to annotate BC findings in clinical cases by using either the pie-star menus, or an array of icons located at the top of the user interface. The array icons and the star menu icons are exactly the same in terms of scheme, color and size. The array allows the sequential selection of finding characteristics, according to the following order: lesion type, type-related characteristics (each characteristic has to

Fig. 3. Array of icons. Lesion type (top). Shape of a mass (bottom).

be described one at a time), comparison to previous studies and location. Fig. 3 presents the array of icons displayed to describe the lesion type (top) and the shape of any mass (bottom).

Nine volunteers ranged in age from 43 to 58 and including 3 females and 6 males participated in this study. All participants were experimented breast radiologists practicing in different hospitals in Belgium. They were recruited regarding their experience in breast cancer screening. Computer skills were assessed thanks to a background questionnaire. All participants were familiar with computers and especially with medical computer-based applications and all were experienced in visual search and navigation activities on computer displays. They were also standard mouse and keypad users with similar quick motor reactions.

The usability test employed a 2x5 factorial design, with two experimental conditions (pie-star menu versus array) and five medical cases to characterize (i.e., five tasks). Each participant carried out ten tasks: five per experimental condition. The order of the experimental conditions was counterbalanced between participants according to a 2x2 Latin Square design. Likewise, the order of the five medical cases per condition was randomized. Counterbalancing and randomization were used in order to neutralize possible task learning effects and to control inter-individual diversity.

The tests were carried out in an isolated room in each hospital. Participants were seated approximately 40 cm from the graphics tablet. Pen-based annotation was used as input modalities whereas visual display was used as output modality. The computer system used in this study was a computer with an Intel Core2 Duo E8400 (3GHz) processor, 4 GB of DDR SDRAM and a 9600GT Nvidia graphic card. The screen was a WACOM CINTIQ 21UX. DICOM (Digital Imaging and Communications in Medicine) images were loaded into the viewer.

The test sessions involved one volunteer at a time. First, participants were given an oral presentation of the project, an explanation of their role in the usability tests, and a demonstration of the functionalities of the tool. Then, they started the training session: one clinical case to annotate per experimental condition. Once they felt comfortable enough with the tool and got used to the manipulation of the pen, they were provided with the paper printed instructions, and the demographic and background questionnaires to fill prior to the effective test. After each condition, they were asked to fill a satisfaction questionnaire. After the two conditions, they were asked to fill the CSUQ [7], a 19-item questionnaire which aims at evaluating the usability of a system in terms of System Usefulness (SysUse), Information Quality (InfoQual) and Interface Quality (IntQual) on a 7-point Linkert scale. A debriefing ended the session. The effective tests lasted approximately 30 minutes.

6 Results

6.1 Statistical Analysis

The sample includes 219 entries (i.e., 219 findings were annotated in all). Analyses of variance (ANOVA) were used to examine the presence of significant differences in task performance, as measured by both annotation times in seconds and number of clicks: per conditions (pie-star and array), per view (CC[1] and MLO[2]), and per finding type (mass, calcification, architectural distortion, special case and associated finding).

Table 3. ANOVA Procedure. Factors: experimental condition, view, and finding type. DF stands for degree of freedom, AT for annotation time, and NC for number of clicks. The statistically differences are bold.

Factors	DF	AT (sec)	NC
Condition	1	F=3.5605; p=0.0605	F=0.0216; p=0.8832
View	1	F=5.6496; **p=0.0183**	F=1.8155; p=0.1792
Finding type	4	F=2.7884; **p=0.0274**	F=9.0073; p<**0.0001**

Table 4. Means and standard deviations of annotation times in seconds

		N	Mean (sec)	Standard deviation
Condition	Pie-Star	115	17.3478	12.8070
	Array	105	20.5524	12.3304
View	CC	138	20.5000	1.0629
	MLO	81	16.3457	1.3874
Finding type	Mass	69	16.2609	1.4949
	Calcification	89	20.9438	1.3162
	Arch. Dist.	39	17.6410	1.9883
	Special Case	5	31.6000	5.5531
	Associated finding	17	18.8824	3.0116

Annotation times. The results presented in the Table 3 show no statistically significant difference between the experimental conditions, but a tendency (F=3.5605; p=0.0605). On the other hand, they show a significant view effect (F=5.6496; p=0.0183) and a significant finding type effect (F=2.7884; p=0.0274).

Number of clicks. The results presented in the Table 3 show a highly significant finding type effect (F=9.0073; p<0.0001).

The results from Table 4 show that the annotation of BC findings with the pie-star menus is faster than with the array of icons (pie-star: 17.5 sec *versus* array: 20.5 sec). Since the results highlight no significant effect of the number of clicks between the two conditions, this difference of speed may be a matter of visualization and visual perception: the star menu displays simultaneously all the icons related to a finding type, whereas the array displays the icons related to one characteristic at a time. In

[1] The Cranial-Caudal view (CC) is taken from above.
[2] The MedioLateral Oblique view (MLO) is taken from an oblique view.

opposition to the array, the star menu enables the users to anticipate their next clicks and, consequently, to be faster.

In addition, the results from Table 4 show that the annotation of BC findings is faster in the CC view than in the MLO view (CC: 20.5 sec *versus* MLO: 16.34 sec). This difference may be explained by the combination of the following two reasons. First, in practice, breast radiologists start the diagnostic by the analysis and interpretation of the CC view. It may be natural to adopt the same task order with the interactive tool. Second, findings such as masses and calcifications need to be characterized in both views. Thus, a "duplicate" button was implemented in order to reduce the number of clicks necessary for the complete annotation of findings.

Table 5. Means and standard deviations of number of clicks

Finding type	N	Mean	Standard deviation
Mass	69	5.46377	0.21331
Calcification	89	4.58427	0.18782
Arch. Dist.	39	4.02564	0.28373
Special Case	5	4.00000	0.79241
Associated finding	17	6.58824	0.42975

Finally, the results from Table 3, Table 4 and Table 5 show that the speed and the number of clicks to perform the annotation task depend on the type of the finding under annotation. This difference may be explained by the combination of the following two reasons. First, the number of characteristics differs from a finding to another (i.e., six characteristics for calcifications, five for masses, and only three for architectural distortions, special cases and associated findings). Second, the important amount of icons to memorize (i.e., about 150) necessarily involves a considerable learning time, and it sounds acceptable that the annotation of unusual findings such as special cases and associated findings requires more time and more clicks in comparison with masses, calcification and architectural distortion which are more frequent.

6.2 User Satisfaction and Preferences

Through the questionnaires and during the interviews, participants considered the interaction with the system as natural, intuitive and reliable. A majority of participants (8) hesitated less than five times, and all participants were satisfied with the compliance with the BI-RADS. Five participants expressed very positive judgments on the star menu in terms of information visualization, speed and comfort. They preferred the star menu because: "it enables the parallel visualization of the items thanks to its spatial organization", "it is more comfortable thanks to its position close to the center of the screen", and "it is faster (than the array)". Four participants preferred the array because "it is usual" and "the characteristics follow a logical sequence".

These results are consistent with the CSUQ results reported in the Table 6. The interface quality, especially, is assessed very positively by the participants.

Table 6. Summary of the overall sample CSUQ. Each 19 item was score on a 7-point Linkert scale (1=totally disagree, 7=totally agree). Statistical indices are mean and standard deviation.

	Mean	Standard deviation
SYSUSE	5.46	0.96
INFOQUAL	5.56	1.11
INTERQUAL	5.81	0.88
OVERALL	5.56	1.89

7 Conclusion

A pen-based interactive tool for standardized annotation of BC lesions was designed and implemented combining user-centered design and usability development methods. Our approach employed knowledge elicitation interviews, scenario-focused questionnaires, paper mock-ups, lab tests, field tests and post-test questionnaires. Additional complementary solutions were designed and implemented in order to facilitate and improve the manipulation of data during the BC finding annotation: the pie and the star menus.

The emphasis of this approach is the attention paid to users and usability. The benefit of this approach is improved user satisfaction. The pie and the star menus lead to better user performances than with the array of icons. This is remarkable with respect to the fact that this unusual interaction style is brand-new and the users thus had no previous experience with it. Furthermore, participants to the usability tests expressed very positive judgments on the star menu in terms of information visualization, speed and comfort and on the user interface. In particular, the interface is judged easy-to-use and adapted to the human activity.

However, as it requires a substantial amount of collected data and numerous individual interviews in great details, the availability of representative users can be an obstacle to the implementation of this novel interaction concept introduced here for other applications.

Acknowledgments. The research presented in this paper has been financed by the Walloon Region within the WIST2-616446 VIGILE project. The authors would like to thank the radiology departments of the Cliniques universitaires UCL (Mont-Godinne), the Institut Jules Bordet (Brussels), the Clinique Saint-Pierre (Ottignies), the Clinique Sainte-Anne (Anderlecht), and the Clinique Notre-Dame (Tournai).

References

1. American College of Radiology, Breast imaging reporting and data system (BI-RADS®) atlas, 4th edn. American College of Radiology, Reston (2007)
2. Callahan, J., Hopkins, D., Weiser, M., Shneiderman, B.: An empirical comparison of pie vs. linear menus. In: Proc. ACM CHI Conference on Human Factors in Computing Systems, pp. 95–100 (1988)

3. Costabile, M.F., Fogli, D., Fresta, G., Mussio, P., Piccinno, A.: Computer environments for improving end-user accessibility. In: Carbonell, N., Stephanidis, C. (eds.) UI4ALL 2002. LNCS, vol. 2615, pp. 129–140. Springer, Heidelberg (2003)
4. Gemo, M., Gouze, A., Debande, B., Grivegnée, A., Macq, B.: A versatile knowledge-based clinical imaging annotation system for breast cancer screening. In: Proc. SPIE Medical Imaging, pp. 6514–6565 (2007)
5. Gouze, A., Kieffer, S., Van Brussel, C., Moncarey, R., Macq, B.: Interactive breast cancer segmentation based on relevance feedback: from user-centered design to evaluation. In: SPIE Medical Imaging, vol. 7260, p. 726021 (2009)
6. ISO 9241, Ergonomics requirements for office work with visual display terminals (VDTs) (1998)
7. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction 7(1), 57–78 (1995)
8. Moyle, M., Cockburn, A.: Analysing Mouse and Pen Flick Gestures. In: Proc. SIGCHI-NZ Symposium On Computer-Human Interaction, pp. 266–267 (2002)
9. Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers, San Francisco (1994)
10. Sefelin, R., Tscheligi, M., Giller, V.: Paper prototyping – what is it good for? A comparison of paper- and computer-based low-fidelity prototyping. In: Proc. CHI 2003 Extended abstracts on human factors in computing systems, pp. 778–779 (2003)
11. Simonin, J., Kieffer, S., Carbonell, N.: Effects of display layout on gaze activity during visual search. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1054–1058. Springer, Heidelberg (2005)
12. Wang, D., Dai, G., Wang, H., Chiu, S.C.: Scenario-focused development method for a pen-based user interface: model and applications. The Journal of Supercomputing (Online First)
13. Wittenberg, T., Elter, M., Schulz-Wendtland, R.: Bildverarbeitung für die Medizin 2007. In: Complete digital iconic and textual annotation for mammography, pp. 91–95. Springer, Heidelberg (2007)
14. Zheng, Y., Wu, M., Cole, E., Pisano, E.D.: Online annotation tool for digital mammography. Academic Radiology 11(5), 566–572 (2004)