UI Prototypes

Identification of software requirements in a highly complex domain

Master Thesis
Date: 09.06.2016
Supervisor: Jan Stage

is109f16

Adam Viggo Glistrup

Dennis Dalgaard

Title:
UI Prototypes - Identification of software requirements in a highly complex domain

Aalborg University
Department of Computer Science
Selma Lagerlöfs Vej 300
9220 Aalborg Øst
http://cs.aau.dk

Topic:
Information Systems

Project period:
Master Thesis
1 February – 9 June 2016

Group:
is109f16

Students:
Adam Viggo Glistrup
Dennis Dalgaard

Supervisor:
Jan Stage

Synopsis:

This report presents the work of two semesters in collaboration with Aalborg University hospital. It is centred around identification of requirements for arterial blood gas results through an IT application. The report includes two articles. The first focuses on working with such a complex domain within limited time scope. A complex domain introduces difficulties in the communication and it can be difficult to identify the requirements for both us and the physicians. The article introduces prototypes as boundary objects in order to facilitate communicate between the two stakeholders. We found boundary objects to externalise tacit knowledge and effective the communication. The use of multiple prototypes enables higher quality feedback and not using prototypes reduces the quality significant. The other articles focuses on the visualisation of rich data through a user interface. We developed an application that was evaluated with eight physicians at the hospital. The evaluation proved that sequential display of results on graphs helped identify trends and contextual information is necessary such as reference interval and treatment information. The adaptation of a new system was difficult which requires time to adjust as well as a customisable interface to fits the specific physician.

Appendixes: 2
Published: 9.06.2016

Preface

This report is the product of group is109f16 and it documents the Master's Thesis in Informatics at Aalborg University. It has been produced as part of the Information System research unit at the Department of Computer Science. The project was done in the timespan from the start of September 2015 to early June 2016. The report contains two articles following CHI conference publication format (ACM SIGCHI, 2016), that are available in the appendix. The report contains 4 chapters. Chapter 1 is an introduction to the project and it presents our problem statement and research questions that serve as the foundation for the articles. Chapter 2 describes the individual contributions of each articles. Chapter 3 describes the research methods used in each article, its strengths and weaknesses and how they relates to the articles. Chapter 4 offers conclusions from each article, a conclusion to the overall problem statement, lists the limitations of the project and presents relevant areas for future work.

Throughout this project, we have had a close collaboration with several departments from Aalborg University hospital. An absolutely vital collaboration for the success of this project. We would therefore like to acknowledge this collaboration by thanking the following people for their contribution to this project:

Biochemist, Claus Gyrup Nielsen, Department of Clinical Biochemistry
Primary contact and intermediary to the hospital.

PhD-student, Olav Lilleholt Schjørring, Department of Anaesthesia and Intensive Care Medicine
Contact at Department of Intensive Care Medicine and intermediary to physicians.

Professor Bodil Steen Rasmussen, Department of Anaesthesia and Intensive Care Medicine
Contact at Department of Intensive Care Medicine.

Leading Cheif Physician, Annebirthe Bo Hansen, Department of Clinical Biochemistry
Intermediary between University faculty and hospital.

Anne Marie Arre, Department of Clinical Biochemistry
Professional insight in blood samples.

Department physician, Jan Nybo, Department of Clinical Biochemistry
Representative of general practitioners.

Chief Physician, Henrik Gregersen, Department of Hematology
Contact at Department of Hematology.

Associate Professor, Louise Bilenberg Pape-Haugaard, Department of Health Science and Technology
Healthcare sparring and insight.

A combined thanks to all the physicians who participated in the evaluations of our application and provided us with valuable feedback and insights.

In addition, we would like to extend a special thanks to our supervisor, Jan Stage, for continuous sparring, discussion and feedback during the entire project.

Abstract

This report covers a nine month long collaboration with several departments from Aalborg University hospital. The collaboration revolved around the development of a software application for visualisation of large amounts of patient data, specifically arterial blood gas results, that is currently only available in paper format. The activities of the collaboration is divided into two articles. This report will briefly present the research questions, research method and conclusion of each article. The full articles can be found as appendix of this report.

The first article focuses on the identifying requirements and all the complication that is included with working with complex domain such as intensive healthcare. The article offers answers on how prototypes can be utilised as boundary objects in order to facilitate better communication for the requirement identification process. Through an explorative case study, we investigate how symmetry of ignorance influence the communication between the involved parties and what can be done to reduce this phenomenon. The study included six meetings and uses seven prototypes as boundary objects. We present discoveries on how and why boundary objects helped to achieve a better communication and create a shared understanding between the involved parties.

The second article focuses on the user interface for visualising arterial blood gas results and how it can support the understanding for highly specialised physicians. We conducted a series of field experiments to evaluate the effect and applicability of the application. The evaluations involved eight physicians from Aalborg University hospital, during which they had to solve two tasks using either the developed application or the current paper-based method. Our conclusion presents different aspect of data visualisation that can improve or worsen the interpretation.

Table of Contents

Introduction 1

Data visualisation is and term that describes visual communication of different types of data. The term conveys an idea that includes more than just graphs and displays of complicated data. *"The information behind the data should also be revealed in a good display; the graphic should aid readers or viewers in seeing the structure in the data."* (Urwin, 2008). When looking at data visualisation as a method for visualising specific data, it is divided into two categories; *exploration* and *explanation* (Iliinsky & Steele, 2011).

A software development process is often a collaboration between developers and intended users of the software in question. When the software is intended for users of a complex domain e.g. healthcare, then this collaboration becomes increasingly complicated. One reason for this is that it involves two very different domains. The communication between designers and users can get clouded by *symmetry of ignorance*, which describes a phenomenon that can occur between designers and users during a design process. It occurs when the two parties are unable to reach a shared understanding of a certain problem, due to the complicated knowledge that needs to be shared (Fisher, 2000).

One of the primary objectives in software development is to identify the requirements of a system. "*A requirement is a statement about an intended product that specifies what it should do or how it should perform*" (Rogers, Sharp & Preece, 2012). It is typically during such a process that symmetry of ignorance can have big consequences. To mitigate this phenomenon, *boundary objects* can be involved in the process. "*Boundary objects are objects that serve to communicate and coordinate the perspectives of various constituencies*" (Arias & Fisher, 2000). These boundary objects are included to function as intermediary between stakeholders and facilitate communication. The boundary objects serves as tools to reduces the symmetry of ignorance and ensure a fruitful collaboration between domains where mutual understanding and comprehensive communication is difficult to achieve.

1.1 Problem Statement and Research Questions

Our master thesis' overall problem statement is as follows:

Problem Statement: *How can we identify software requirements through prototypes within a highly complex domain and develop a user interface that supports the understanding of that domain by visualising rich data?*

The objective of this thesis is the study of software requirements, prototypes, complex domain, user interface and rich data. Software requirements are increasingly difficult to identify as the complexity of a domain increases. We set out to investigate how user interface oriented prototypes can help with this issue. A highly complex domain combined with rich data introduces the problem of how to visualise the data. We wanted to investigate how to improve user interfaces to support this understanding. To answer our problem statement we split it into two separate research questions:

Research Question 1: *How can multiple prototypes be used as boundary objects to identify software requirements within a highly complex domain with limited resources for collaboration thus reducing the symmetry of ignorance?*

The first research question investigates how prototypes can be utilised as boundary objects to identify software requirements. The context for this research question was set in a highly complex domain with limited resources. High complexity often requires more time and money to manage, and the research question was oriented towards reducing this problem. The complexity between two parties is also known as symmetry of ignorance which is the research question's goal to reduce.

Research Question 2: *How can a user interface support highly specialised domain experts' understanding by visualising rich data?*

The second research question investigates how to visualise rich data for highly specialised domain experts. In extension of the first research question, the rich data also consists of a problem as the meaning of the data is obscure and requires the expertise of domain experts. The research question focuses on how a user interface can support the understanding of domain experts by providing the right context and type of visualisation.

1.2 Case Description

In order to investigate the problem statement of this project we collaborated with various departments from Aalborg University hospital. A collaboration between IT students with a specialisation in HCI and highly specialised medical physicians; two radically different domains. The collaboration was centred around the development of an application for visualisation of arterial blood gas results. Arterial blood gas samples is used whenever a patient is critically ill, "*it allows practitioners to assess the adequacy of ventilation, oxygen delivery to the tissues and acid-based balance status*" (Simpson, 2004). The results from analysing arterial blood gas samples is considered rich data, because a single blood sample contains between 20 and 30 components and it can be necessary to analyse a single patient up to 30 times a day.

The visualisation application is meant to support and aid physicians and nurses in the diagnostic and monitory process of patient at an intensive care unit. A process that is regarded as very individual and can be approached maybe different ways e.g. *stepwise* or *systematic* (Barnette & Kautz, 2013). The focus of the application was to visualise result data, provide a clear view of trends over time and enable quick identification of radical fluctuation. This required combining elements from data visualisation, medical diagnostics procedures and the technical development.

The application was evaluated to determine its practical performance. This evaluation was conducted as field experiments, where eight physicians was asked to perform certain diagnostical tasks using both the current paper method and the developed application.

1.3 Research Process

The work presented in this report is a continuation of work done as part of a 9^{th} semester project. During which we investigated the impact of prototypes as focal point for communication between developers and users. The results from this previous work serve as partial foundation for the work with research question 1 in this project.

Contributions 2

In this chapter, we present the two articles of this project, which represents the main part of this master thesis. A short overview and description of the relationships between them will be presented, followed by a short summary of each article and its contributions. A chronological reading of these articles is highly recommended as their individual content represent a part of a larger process and is beneficial to read in the right sequence. However, each article can be read individually as they tackle different issues, despite the context being the same project.

2.1 Contribution 1

Glistrup, A. V., & Dalgaard, D. (2016). *Utilising boundary objects to identify software requirements in a highly complex domain: An explorative case study of symmetry of ignorance*. Aalborg: Department of Computer Science, Aalborg University.

This article contains an explorative case study of the phenomenon symmetry of ignorance (Fisher, 2000). The case study revolved around a collaboration between Aalborg University hospital and the authors, during which symmetry of ignorance occurred. The goal of the study was to identify methods for reducing this phenomenon.

In order to reducing the symmetry of ignorance, we introduction boundary objects (Arias & Fischer, 2000) for the meetings between stakeholders. The boundary objects of this study were a series of simple and functional prototypes, showing a variety of different options and possibilities. The goal was to enable the both parties in a more engaging and fruitful conversation. We have analysed the physicians' means of expression with different prototypes during six meetings. These meetings took place Aalborg University hospital and were conducted as *semi-structured interview* (Kvale, 1997).

The meetings revealed that the use of prototypes helped communicate tacit knowledge. Instead of discussing what they thought they needed, prototypes was able to externalise tacit medical understanding into a common language. The prototypes was effective to generate feedback and resulted in very little time required for collaboration. We also found that using multiple prototypes increases quality as the feedback was more concrete, while not using prototypes reduces the quality as the feedback was more vague and indecisive.

2.2 Contribution 2

Glistrup, A. V., & Dalgaard, D. (2016). *How can a user interface support understanding of rich data? Evaluation of an application for visualisation of arterial blood gas.* Aalborg: Department of Computer Science, Aalborg University.

This article contains a field experiment of a developed application based on the results of the first contribution. The application attempts to redefine and improve how physicians at the Intensive Care Unit of Aalborg University hospital can interpret rich patient date, specifically arterial blood gas results, through data visualisation. The article presents an evaluation of the application and compare it to the current practice.

In order to evaluate the application, we enlisted eight physicians from three departments at Aalborg University hospital, all with connection to the Intensive Care Unit and experience with analysing arterial blood gas results. The physicians were asked to analyse two patients based on a short patient history and arterial blood gas results. The results was presented using either the application or the current method, which is paper-based observation forms. After analysing the patients, the physicians was asked to evaluate their analysis and performance on a questionnaire followed by a *semi-structured interview* (Kvale, 1997). This created both quantitative and qualitative data that lead to the results of this article.

The results revealed that the application had several benefits for improving interpretation despite the physicians' lack of experience with the application. The improvements were the display of trends, reduced time to see tendency, new discoveries, context of the results, and useful for reviewing new patients.

The overall findings of the study revealed very positive feedback on visualising arterial blood gas results. However, minor changes to the application would be required in order to accommodate the many different approaches to diagnosing that were identified, as well as the inclusion of additional information currently only stored on the paper observation forms. The large amount of different approaches to the same problem incentivises the use of adaptive or customised user interfaces. The application did not provide the same detailed overview of single blood samples which requires either changes in the application or time to adjust to the new visualisation.

Research Method 3

This section describes the research methods used in the articles and reflects over their strengths and weaknesses.

3.1 Research Method 1:

The first article can be classified as a case study which is described to be used for explanation, description and hypothesis developments (Wynekoop & Conger, 1990). In the research we sought explanations and descriptions of the phenomenon, symmetry of ignorance, we were investigating and developed our hypothesis in regards to how to use boundary objects in our collaboration. The research method can be further defined as an exploration case study as it "*begin with an incomplete or preliminary understanding of a problem and its context*" (Lazar, Feng, & Hochheiser, 2010, s. 150). We have no predefined solution or fixed process but are all subject to change while we undergo the exploration of this case study. The drawbacks addressed with case studies is its high cost, time and limited generalisation (Wynekoop & Conger, 1990). High cost and time are negated due to the nature of irrelevant costs as students and time scoped by deadlines. To countermeasure the lack of generalisation, we use two different departments to increase reliability.

3.2 Research Method 2:

The research method can be considered a field experiment which is "*experimental manipulation of one or more independent variables while contaminating variables are controlled to observe the effect on dependent variables, all in a natural settings*" (Wynekoop & Conger, 1990). The advantages of having control while still remaining in a natural settings is the increased the reliability. The disadvantage is if the control becomes to dominants, it may reduce the reliability (Wynekoop & Conger, 1990). To negate this disadvantage we controlled as few variables as possible such as the information available of the patient, the tools available for the analysis and the time available. The uncontrolled variables were the procedure and the medical content produced by the participants. Of these variables we only used the produce for our research and considered the medical content irrelevant for the research but necessarily to maintain natural settings. To further increase the natural settings all evaluations was conducted at the participants' local office during normal work hours. Another disadvantage is its difficulty to collect enough data, which is countered by our collaboration with Aalborg University hospital that enabled us to evaluate with eight physicians.

Conclusion 4

Following section summarises the conclusion for the two research questions and the overall problem statement followed by limitations and future work.

4.1 Research Question 1

How can multiple prototypes be used as boundary objects to identify software requirements within a highly complex domain with limited resources for collaboration thus reducing the symmetry of ignorance?

Prototypes can be used as boundary objects to improve communication of tacit understanding. It externalises implicit thoughts and assumptions into a common language and context. It also enabled more effective feedback, which is great for time limited collaborations. The feedback derived with the prototypes were concrete and more easily translated into requirements that we understand. Not using prototypes resulted in more vague and indecisive feedback. Multiple prototypes provided even better feedback as the relation between the prototypes creates a more deliberate consideration of the different aspects.

4.2 Research Question 2

How can a user interface support highly specialised domain experts' understanding by visualising rich data?

We found trends to be an essential part of the visualisation to see whether values are rising or falling over time. It provided a greater understanding than current practice and enabled new conclusions for what treatment should be conducted. It is also important to gain an immediate understanding of the current status: We found the reference interval to be a good example. However, the presented application lacked proper graph sizing as all results were too huge to gain the necessary information fast enough. Another aspect we found that can help set the context was to associate the treatment with the data. The domain experts were hugely different and have different requirements hence an identical application for all physicians is not recommended. The domain experts must to a great extend to able to adjust their personal preference in the visualisation such as what to be available, which order, and how it appears.

4.3 Problem Statement

How can we identify software requirements through prototypes within a highly complex domain and develop a user interface that supports the understanding of that domain by visualising rich data?

Using prototypes as boundary objects provided a common understanding between IT domain and medical domain in regards to the software requirements. This approach aimed to communicate about a solution rather than the process of reaching a solution. The solution takes the shape of a user interface which serves as an interface between the stakeholders as it both gives IT domain and medical domain an application reference. A user interface can support rich data by visualising the results in a context such as trends, reference interval, and treatment information. As the understanding of complex domain is largely unknown, it is important that the visualisation is adjustable to the specific user to counter the uncertainty associated with a highly complex domain.

4.4 Limitations

Bias

The research we present is both described and conducted by ourselves. This raises the issue of whether we are able to remain objective to the result as we are biased towards the findings. Our perspective may differs from others' perspective and it is uncertain if the conclusions reached would be the same for other researchers. Other researchers may interpret methods and approaches differently and ends up with different results. As our use of the methods may assume certain perspective of interpretation that are not eliminated by other peoples' use of the methods.

Subjective Evaluation

A limitation with the evaluations is the approach to conclude based on the physicians' opinions and decisions. By using their subjective assessments we introduce another aspect of possible misleading communication. The physicians' perspective includes the process of communicating about their conscious thoughts which is a source of false reasoning and misinterpretation: Physicians, like any other users, are most likely not aware of their actual needs but can only a reflect from previous experience and current practice.

Short Term Evaluation

Another limitation with the evaluations is the duration. The settings for the experiment is an unrealistic short time to draw a fair comparison between their current practice (observation form) from the new solution (application). It takes time for users to learn and adjust their habit and routines to other applications. Although a brief introduction for the application was given, it does not fully counter the extensive experience they have with the observation form.

Limited Generalisation

Even though the scope of the project is a rather concrete problem for the medical practitioners, it is difficult to prove the generalisation reaches that abstraction level. Both contributions deals with a specific case and does not replicate the method much. The first contribution uses two meetings with another department and the second contribution uses eight physicians in the evaluation. However, this does not fully justify the findings to cover the scope of a highly complex domain, rich data and highly specialised domain experts.

4.5 Future Work

Improve Visualisation

Our findings in the second contribution highlights which elements an arterial blood gas visualisations should have. However, it also raises some issues with the developed application such as lack of treatment information, the complexities of the different components, the missing immediate understanding, and the very different usage. The obvious future work would be to eliminate these issues in order to construct an even better application. This could add value the medical sector which is the core reason why this research is relevant for them.

Long Term Evaluation

As mentioned, the short term evaluation is a disadvantage to our understanding. It would be relevant to investigate how the application performs in the long run. This is quite a common problem for all freshly developed systems and would be interesting for the IT field to investigate as a general problem. If we were able to negate some of the drawbacks to conduct a more fair comparison it would provide reliability to the evaluation.

Increase Generalisation

As methods in both contribution is applicate on other instances, it would be relevant to redo the experiments with other hospital and other physicians; "*Replicated case studies may be used to increase generalise results*" (Wynekoop & Conger, 1990). By replicating the experiments and comparing the findings it would be more conclusive rather than suggestive whether our findings is sufficiently abstract.

Solutions as learning

Identifying software requirements can be considered a learning process as you learn the requirements through some kind of identification process. We found that solutions, such as prototypes, works great in this process. It would be interesting look at other fields to investigate how solutions can help learning. Instead of learning the fundamentals of a complexity, maybe you need the solution in order to understand the fundamentals. This question would be relevant to schools and any other learning instances.

References

ACM SIGCHI. (2016, May 29). *CHI 2016.* Retrieved from Guide to Submission Formats: https://chi2016.acm.org/wp/guide-to-submission-formats/

Arias, E. G., & Fischer, G. (2000). Boundary Objects: Their Role in Articulating the Task at Hand and Making Information Relevant to It. *Intelligent Systems and Applications*, pp. 1-8.

Barnette, L., & Kautz, D. D. (2013). *Creative Ways to Teach Arterial Blood Gas Interpretation.* Lippincott, Williams & Wilkins.

Fisher, G. (2000). Symmetry of ignorance, social creativity, and meta-design. *Knowledge-Based Systems*(13), 527-537.

Iliinsky, N., & Steele, J. (2011). *Designing Data Visualizations.* Sebastopol: O'Reilly Media, Inc.

Kvale, S. (1997). *Interview - En introduktion til det kvalitative forskningsinterview.* København: Hans Reitzels Forlag.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction.* West Sussex: John Wiley & Sons Ltd.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human-computer Interaction.* West Sussex: U.K. Wiley.

Simpson, H. (2004). Interpretation of arterial blood gases: a clinical guide for nurses. *British Journal of Nursing, 13*(9), 522-528.

Urwin, A. (2008). Good Graphics? In C.-h. Chen, W. Härdle, & A. Unwin, *Handbook of Data Visualization* (pp. 57-78). Berlin: Springer.

Wynekoop, J. L., & Conger, S. A. (1990). A Review of Computer Aided Software Engineering Research Methods. In *The Information Systems Research Areas of the 90's: Challenges, Perceptions, and Alternative Approaches* (pp. 130-154). Lund.

Appendixes

Appendix 1:

Glistrup, A. V., & Dalgaard, D. (2016). *Utilising boundary objects to identify software requirements in a highly complex domain: An explorative case study of symmetry of ignorance*. Aalborg: Department of Computer Science, Aalborg University.

Appendix 2:

Glistrup, A. V., & Dalgaard, D. (2016). *How can a user interface support understanding of rich data? Evaluation of an application for visualisation of arterial blood gas.* Aalborg: Department of Computer Science, Aalborg University.

Utilising boundary objects to identify software requirements in a highly complex domain: An explorative case study of symmetry of ignorance

Adam Viggo Glistrup
Aalborg University
Department of Computer Science
Aalborg, Denmark
adam@adamglistrup.com

Dennis Dalgaard
Aalborg University
Department of Computer Science
Aalborg, Denmark
ddalgaard@gmail.com

ABSTRACT
Identifying software requirements within a complex domain is difficult with limited time available. This requires close collaboration between stakeholders. However, when working with a complex domain this collaboration can be clouded by communication difficulties. This problem is known as symmetry of ignorance. To investigate this phenomenon, we have collaborated with a highly complex domain, represented by Aalborg University hospital, to identify software requirement using boundary objects. The collaboration consisted of six meetings and included a total of seven prototypes that served as boundary objects. We discovered using boundary objects helped to identify software requirements by externalising tacit knowledge. Boundary objects enabled effective feedback with the limited time available. Additionally, we discovered that using multiple prototypes provides feedback of better quality than a single prototype. Oppositely, not using any boundary objects was found to reduce the quality of the feedback as it was vague and indecisive and therefore less usable for identifying requirements.

Author Keywords
Symmetry of ignorance; boundary objects; prototypes; requirements; complex domain

INTRODUCTION
Software development is a vast and ever expanding field of different methods and procedures with different advantages and drawbacks depending on the given situation. The well-known procedures include traditional waterfall model [1] and various types of agile development [2]. However, when working with highly complex fields such as healthcare, that involves a very versatile working environment, then these common procedures are not always well suited.

Back in 1984 Horst Rittel coined the phrase *Symmetry of Ignorance* [3] which describes a phenomenon that occurs between the designers and users, when a complex domain requires expert knowledge from both parties but they are unable to reach a shared understanding, due to the complexity of the knowledges [3]. This communication barrier is increased when designers are unable to emerge themselves in a given problem due to limited resources such as time restriction. Furthermore, the users are often unable to fully communicate their needs. This is often caused by complexity of the problem and as a result of the users not knowing what they actually need.

Several practitioners of HCI have suggested the use of boundary objects [4][5][6] as a technique to facilitate communication when the gap between designers and users become too wide. These boundary objects can take shape of physical objects with different representation, drawing, illustrations, and more. Examples of this is shown on figure 1. The common purpose for all the representations is to create a shared understanding and coordinate the perspectives of various constituents [6].

Figure 1. Examples of boundary objects in use [6][7].

Highly complex domains produces complex problems. Complex problems is defined as requiring more knowledge than any single person possesses [4], so clear and explicit communication through collaboration between stakeholders is essential for a successful solution. According to the the Cynefin Framework [8] a complex domain can be described as: ”*A domain where the relationship between cause and effect is unclear*”[9]. In such a case the framework dictates to conduct experiments in order to identify what works.

The aim of this paper is to investigate how multiple prototypes can be used as boundary objects to identify software requirements within a highly complex domain with limited resources for collaboration thus reducing the symmetry of ignorance. This is done through a project in collaboration with Aalborg University hospital, where the goal is to develop an application that can assist physicians by visualising large amounts of patient data. The understanding of this application's requirements serve as the foundation for this paper.

In the following sections we present related papers in regards to understanding the term symmetry of ignorance and working with prototypes to determine requirements. Then our method and collaboration with Aalborg University hospital is presented and described, including the prototypes used in the process. The findings emphasise the outcome of the collaboration and the work with boundary objects. The discussion compare our results to others and discuss the limitations of this study. The paper ends with a conclusion that sums up the essentials of our findings.

RELATED WORK

Other authors have investigated the effects and consequences of symmetry of ignorance in a design process. One of the most influential is Gerhard Fishers conceptual framework for creativity [4] in which he argues that the mutual ignorance of the stakeholders within a given project can be used as positive attribute in the design process. His framework uses boundary objects as an abstraction tool to capture tacit knowledge that leads to an increase in shared understanding [4].

Another approach is to minimise the effects of symmetry of ignorance. One suggestion is described by Shantanu Pai and Kenneth R. Allendorfers [5]. They introduce a boundary object to the design phase that serves as a facilitator for the communication about concept ideas. In this case, the boundary object is an interactive prototype of the system. The prototype undergoes several iterations before reaching the final design. During each iteration, the prototype is used to facilitate new communication and gather more knowledge for understanding of the system's requirements.

In order to comprehend all the aspects of prototypes, we refer to Christies et. al. work on prototyping strategies [10]; a literature review that focuses on the general areas of prototyping. The article is mainly centred around prototypes in industrial engineering as physical prototypes are very common. However, it does provide valuable insight and a conceptual understanding about the creation of software prototypes. The article identifies and lists different characteristics of varied design problems and suggest how to plan a prototyping strategy accordingly. The most relevant concept for this paper is the analytical prototype that is used to analysis and define the requirements in the early design stages [10].

The use of multiple prototypes to define requirements as part of the development process for medical software has been attempted before. One example is Sven Koch, Amanda Sheeren and Nancy Staggers' work from 2009 [11], in which they created four different paper prototypes of a monitoring interface with varied levels of information and interaction. These prototypes was produced to match four different personas. These personas was created based on semi-structured interviews with several nurses that represented potential users of the system. Subsequently, was each prototypes evaluated with nurses in order to gather feedback and to identify if the nurses actually preferred the prototype connected to the persona that best resembled their own individual persona. The article recommends using a broad range of personas in order to gather a diverse perspective of the required functionality.

In regards to this paper, the main goal is to combine the concepts of prototypes and boundary objects in order to address the presence of symmetry of ignorance that often occurs when working with a highly complex domain.

METHOD

We have collaborated with Aalborg University hospital to develop requirements for software representation of patient results. The hospital was represented by physicians that are highly specialised in their profession and their work area can be considered a highly complex domain. They analyse patient results in their daily job and can therefore be considered to be actual end users. The collaboration involved a total of six meetings with the physicians. All meetings was conducted at Aalborg University hospital. The materials brought to the meetings were laptops with the relevant prototypes. Notes were taking during the meetings which were processed through a meaning condensation after the each meeting [12].

We collaborated with two departments: Department of Hematology and Intensive Care Unit. Department of Hematology was represented by a Chief Physician and involved the problem of representing certain components in multiple myeloma (bone marrow cancer). These components and their relation to each other suggests certain medical trends of how well treatment is progressing and how the cancer is maintained. We held two meetings with this Chief Physician over the course of three months. The other department, Intensive Care Unit (ICU), was represented by a Chief Physician and a PhD student who formerly worked as a physician in the department. ICU includes a much larger variety of patient results compared to the case of multiple myeloma and the problem is very different.

Table 1 summarises all the meetings, who participated, the duration of the meeting and what materials was presented during each meeting.

Meeting	Department	Task	Users	Duration	Material
#1	Hematology	Introduction and idea sharing	Chief Physician from Hematology	60 minutes	Generic prototype of graphs (see Figure 2)
#2	Hematology	Feedback on prototype	Chief Physician from Hematology	30 minutes	Multiple Myeloma prototype (see Figure 3)
#3	Intensive Care Unit	Introduction and idea sharing	Chief Physician from ICU	60 minutes	Hematology prototype (see Figure 3)
#4	Intensive Care Unit	Feedback on prototypes	Chief Physician from ICU	90 minutes	Four ICU prototypes (see Figure 4, 5, 6, and 7)
#5	Intensive Care Unit	Scoping of further development	Chief Physician and PhD student from ICU	90 minutes	None
#6	Intensive Care Unit	Feedback on prototype	PhD student from ICU	75 minutes	Single ICU extensive prototype (see Figure 8)

Table 1. Overview of meetings.

FINGINGS

The physicians proved to have a very busy schedule that resulted in limited time available for meetings. This aligns with our intend to investigate the problem in a context where time is considered a valuable resource. Additionally, we confirmed the physicians had very limited understanding of software development and could only reflect on it from normal user's perspective. This was determined as fundamental understanding of technical IT knowledge clearly was lacking. Likewise, did we not have any prior medical knowledge and are completely unaware of any practice in the field of medicine. As both parties are acknowledged ignorant to the other field, the criteria for symmetry of ignorance is fulfilled.

Meeting #1

The initial meeting provided us with an introduction of the medical condition (multiple myeloma), current software usage and his experienced problems. This meeting also gave us an opportunity to present a generic non-case specific software prototype (see Figure *2*) developed prior to the meeting. This prototype showed how data can be represented in different ways without any medical context. The prototype includes a line representation of four different datasets which can be toggled on and off. The data are arbitrary placeholders that emphasise the development over time.

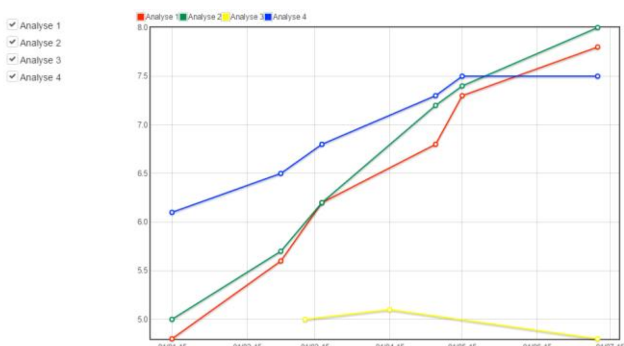

Figure 2. Prototype of generic graph.

It also attempts to signify the relationship between the datasets as it is visually comparable due to being on the same graph.

The prototype allowed the physician to concretise the tacit understanding of his needs from medical terminology to implications in the prototype. As a result of this communication we were able to expand our understanding of the usage requirements and therefore ask more concrete questions that served us better. For instance did the physician determine that exact time of the analysis was not necessary by looking the time axis and continued to explain that the value did not vary depending on the time of the day. This leads to increased understanding for us as well as pushing him to reflect on his usage of the patient data. He commented: "*The graphs certainly does something*" as a positive regard to the visualisation. The physician suggested that this type of graphs would help identify multiple myeloma as it consists of two components, Kappa and Lambda, that are directly related to one another. We discussed this illness and by using the prototype the physician was able to relate the complex medical description to a very specific IT application. He mentioned: "*It is important to see the reference interval with these graphs*". Reference interval is the common value for the average citizen. The physician also showed us their current IT system and complained about the poor usability of the date filter. He was not able to determine what the best solution would be but could only criticise the current practice.

Meeting #2

At the second meeting we presented our next prototype (see Figure *3*). The presentation included thoughts and assumptions made throughout the development of the prototype with the intention of being as transparent as possible. The prototype includes two main graphs: On the upper graph, the components, Kappa and Lambda, are shown in relation to one another similar to figure 2. The middle graph shows the ratio between these components as a single dataset. The white area on the second graph represents the reference interval. As the ratio often reaches very high value, it is possible to toggle between a fixed pre-defined view and an auto-adjusted view. The predefined view locks the maximum value shown while the auto-adjusted view scales with the value of the ratio.

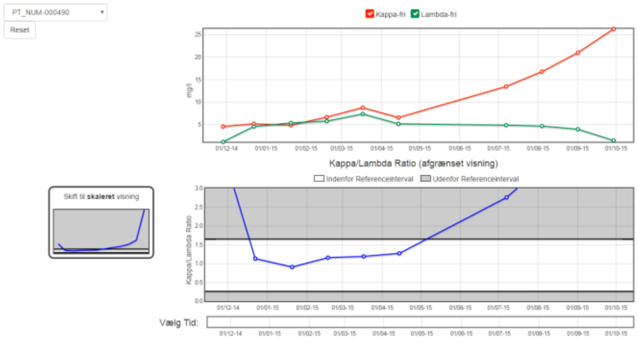

Figure 3. Multiple Myeloma Prototype.

We were able to quickly identify flaws in the previously formed understanding and assumptions. Within the first minute he requested another component that was not shown; the concentration of Kappa and Lambda. He reasoned that this is the largest indicator of how dominant the cancer is. After inspecting the prototype, the physician expressed that the system would help to identify plateau phases (when the treatment can be halted). This usage was not discussed prior to the meeting and reveals that the prototype can invoke expanded understanding for the physicians as well. He also commented: "*The scaled graph is not necessary. If it is way beyond the upper limit it is not important how much is goes beyond*". This allowed us to effectively gain an understanding of the requirements by relating our combined software knowledge and medical expertise to the prototypes. The meetings with the Department of Hematology concluded here and we continue our work with ICU.

Meeting #3

During the first meeting with ICU we set the scope for what kind of result we would be processing: arterial blood gas. Only the Chief Physician was present. She described her problems with their current practice and what she thought they needed. We showed some solutions from our case with multiple myeloma, from which she pointed out elements she wanted and elements she did not want: She liked the graphs but not the focus on the reference interval as "*the reference interval is in reality different for every person and situation and does not always indicate the best for the patient. We are handling very ill patients and often intentionally ignore the reference interval*". She was not able to explain her reasoning for what she thought she wanted: "*Everything should be shown on the graphs*" and "*Nothing is more important than the others. It all depends on the situation*". It was difficult for her to determine what she needed from an IT system and her comments seemed vague and indecisive. When discussing usage for the departments she comments: "*It would be great if there could be an app with this*" and "*For chronical treatment, there should be an algorithm that can recognise the individual patient*". These statements indicated a lack of focus on her actual needs as they are not very specific regarding the requirements. She requested what she believed was the correct answer and not based on her actual needs.

Meeting #4

Prior to this meeting we developed four prototypes based on our experience from the previous meeting and different ideas during the software development. These prototypes showed different ways to visualise a lots of different arterial blood gas components . Prototype A (see Figure 4) shows the latest results for each component, including the value, time, unit, and whether it is above, below or within the reference interval. Clicking on a components reveals a graph of all results for this component.

Figure 4. ICU Prototype A.

Prototype B (see Figure 5) took a different approach and visualises all results of each components as graphs immediately. However, this view does not provide the user with detailed information about reference interval or the latest result.

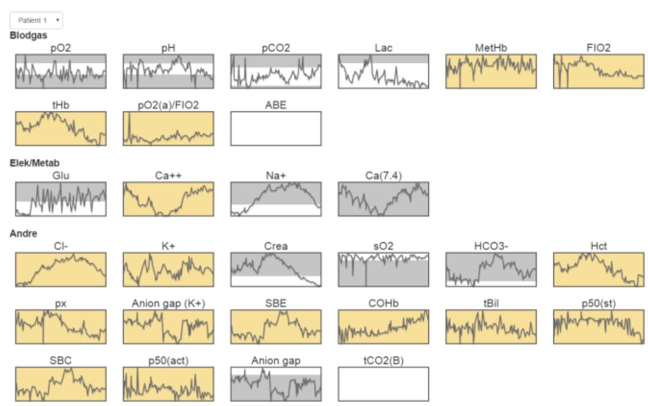

Figure 5. ICU Prototype B.

Prototype C (see Figure 6) combined the two visualisations into a single view at the cost requiring much more space and information conflicts.

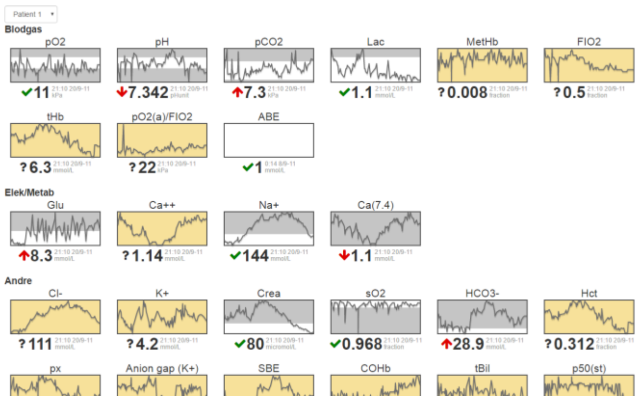

Figure 6. ICU Prototype C.

Prototype D (see Figure 7) visualises the results as a much more detailed view for each components. This requires a lot more scrolling and reduces the immediate understanding of all the results.

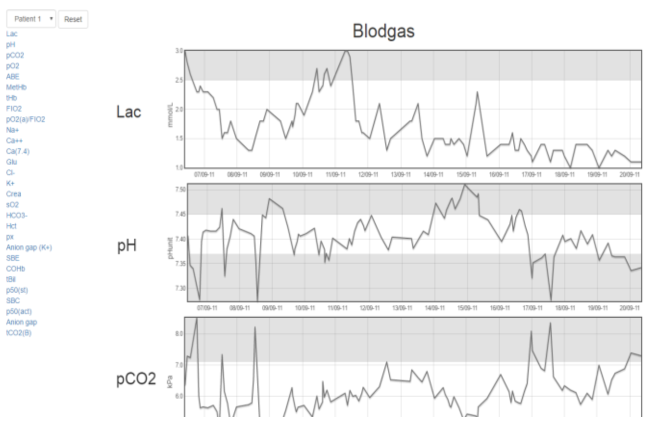

Figure 7. ICU Prototype D

At the meeting we presented the prototypes, individually, to the Chief Physician and discussed what usage and benefits each prototype provided. The presentation of the prototypes broadened the physician's perspective as she began to express new ways the prototypes could benefit their daily work: The graphs can be used to understand treatment courses after the treatment has ended to improve the medical practice. She also concluded "*We are giving way too much oxygen*" by analysing the prototype for herself. This indicates she learned about medical practice through the prototypes. She continued: "*Details are much more important than overview*" which referred to prototype D as being superior to the others. The different prototypes allowed her to reflect on aspects of IT implementation that previously was considered vague and indecisive. She is suggestive of more concrete ideas such as packages for enabling certain components, discussing the practical impact of seeing results over longer time and the how different components view relates to each other. In comparison to meeting #3, we experienced much better feedback that we were able to use as non-medical actors. The feedback was very concrete and directly usable for further development. We notice this significantly more for each prototype we introduced, suggesting that the usage of multiple prototypes provides high quality of the feedback.

Meeting #5

At this meeting both the Chief Physician and the PhD student were present. The meeting's agenda was to clarify some requirements uncertainties and specify the setting for future collaboration. It was also the first time the PhD student was introduced to the project as the Chief Physician had problems to find time for our meetings. No prototype or new software was introduced during this meeting and we experienced much less useable feedback than the other meetings. We discussed the system in a much higher abstraction such as the common usage for the system, what visualisation generally can provide to them and what is important for their department. We spent most of the meeting talking about medical specific problems which was far beyond our level of comprehension but at the same time seemed like basic knowledge for the physicians. This was a significant amount of time spent to gain very little understanding. The meeting was highly medically oriented and attempted to educate us to understand the domain rather than finding a solution. This attempt failed as we gained very little new knowledge and therefore not a much better understanding of the requirements. The vague and indecisive opinions emerged again and reduced the quality of the feedback.

Meeting #6

This meeting involved only the PhD student. A more extensive prototype was developed for this meeting based on our previous experience and interviews with the physicians (see Figure 8). The prototype was based on Prototype D (see Figure 7) with major functionality improvements. The prototype include selection of patient via Center Person Registry (CPR), what time period to show, what patient type to show and a customisable reference interval.

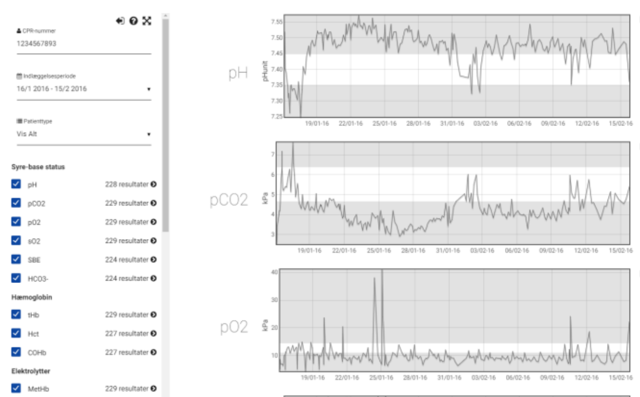

Figure 8. ICU Extensive Prototype.

The PhD student had very concrete comments about how certain elements and categories should be repositioned due to medical reasoning. He was very quick to identify these elements and what exactly should be done. Additionally, he had several suggestions about how some interactions could be different, according to his perspective. For instance, some of the components included too many decimals or the naming convention did not match his perception. We discussed this with him and used our technical expertise to reason about which changes would make sense. This gave us extremely concrete feedback that was easy to translate to software requirements without any difficult interpretations.

Summery

Prototypes helped communicate tacit understanding

Using prototypes as boundary objects allowed us to rapidly communicate complex software solutions. It externalised the physicians tacit understanding of a medical problem into a common language. This was established as the physician from Intensive Care Unit expressed very positive reactions during meeting #4 and #6. Both our and the physicians' understanding of the requirements was increased as a result of using boundary objects.

Prototypes enabled effective feedback

We had a total of 6:45 hours collaboration time with the physicians and still managed to produce well-received boundary objects in form of prototypes. This indicates the prototypes worked effectively to generate feedback and set a precise scope for what we were working with. The context of the boundary objects provided us with a dynamic effect at the meetings; it allowed rapid communication between us and the physicians' to quickly determine the logic of the other parties arguments. This made understanding of the requirements more comprehensible in a very time restricted setting.

Multiple prototypes increases quality of the feedback

We discovered at meeting #4 that the use of multiple prototypes increased the quality of the feedback as it was more concrete. The physician was able to reason the different positive and negative effects and pinpoint exactly how it related to the medical use. The users are able to better perceive the strengths and weaknesses of the different prototypes in relation to each other. This increases the rationale of the arguments thus increases the quality of the feedback.

Not using prototypes reduces the quality of the feedback

At meeting #5 no prototypes was discussed and we experienced that we gained much less usable feedback from this meeting than the other meetings with prototypes. Merely discussing requirements as general instructions without any prototype to provide exact context resulted in vague and indecisive feedback that did not benefit much to our understanding. This was also established during meeting #3 when talking about future features. We experienced that medical terms, descriptions and concepts explained by the physicians was harder to comprehend, and we had to request a lot of elaborations on the subjects. The communication was restrained as our common knowledge between the different fields of expertise was very limited.

DISCUSSION

The scope of this paper is to identify the requirements for a software application but not to cover the complete software development process. Isolating the research to a limited part of the development process can result in an incomplete understanding of the domain. Identifying the requirements is often associated with the early part of a software development process and this paper does not consider the impact it may have beyond that. An argument to improve the usefulness of our findings could be to increase the scope and investigate the long-term implication of using prototypes as boundary objects.

During this study, we only investigated the impact of using multiple prototypes during a single meeting. This raises the question whether it actually was the use of multiple prototypes that caused such great feedback or if the result was caused by other factors. Verifying the effects of multiple prototypes could have been improved by presenting multiple prototypes during several meetings. However, a literature review of prototyping strategies by Christie et al. [10], support this finding. The researchers behind the study state that: "*By creating a large number of scaled prototypes at a low resource cost, creators are able to rapidly conduct user evaluations of the product as well as start understanding trends in user preference at an early stage in the process.*" [10].

Another similar problem occurred at the meeting without prototypes. As the physicians already had seen the previous prototypes, their opinions and feedback might be affected by previous knowledge. Again, we cannot truthfully determine the validity of our findings because the investigated factors were not isolated. This problem is also addressed in Christie's et al. literature review [10] through claims such as: "*Never go into a meeting without a prototype.*" [10].

This research paper is largely based on others work [4] [5] [6], and reach similar conclusion, such as: "*[about boundary objects] help in addressing the issues that design teams face by facilitating cross-domain communication*"[4]. In addition to this, Gerhard Fisher concludes that, by addressing the symmetry of ignorance and encourage communication, the result is a more creative solution [5], as stakeholders are forced to think and act in alternative ways. This result is similar to what we encountered during our meetings. It was clear that one of the physicians was very impressed with the possibilities of the project as it progressed, even though most the majority of the ideas came from her or caused by a joint effort during meetings. In other words, she was surprised by her own creativity, triggered by working with the boundary objects.

CONCLUSION

Using prototypes as boundary object within a highly complex domain helps communicate tacit knowledge and therefore reduces the symmetry of ignorance. This reduces the time required to identify software requirements as the prototypes enables effective feedback between the parties.

Using multiple prototypes results in better quality of the feedback as the comparison between the prototypes enables more concrete and usable feedback. Oppositely, not using prototypes reduces the quality of the feedback as it becomes vague and indecisive.

Further research in reduction of symmetry of ignorance might be necessary in order to validate our findings as they are currently based on single cases and assessments.

ACKNOWLEDGMENTS

We would like to thank Aalborg University hospital for their involvement in the process. Specifically, we would like to thank the participating physicians at Department of Hematology and Intensive Care Unit as well as the assisting people associated with the project. We would also thank Jan Stage for providing supervision throughout the entire project.

REFERENCES

1. L. Mathiassen, A. Munk-Madsen, P. A. Nielsen and J. Stage, Object-oriented analysis & design, Aalborg: Marko, 2000
2. C. Larman, Agile and Iterative Development: A Manager's Guide, Addison-Wesley Professional, 2004
3. H. W. J. Rittel, “Second-generation Design Methods,” in *Developments in Design Methodology*, New York, John Wiley, 1984, pp. 317-327
4. G. Fisher, “Symmetry of ignorance, social creativity, and meta-design,” *Knowledge-Based Systems*, no. 13, pp. 527-537, 2000
5. S. Pai and K. R. Allendoerfer, “Reducing the "Symmetry of Ignorance" Between Human Factors Engineers and Users,” *Ergonomics in Design*, vol. Summer, pp. 24-29, 2006
6. E. G. Arias and G. Fischer, “Boundary Objects: Their Role in Articulating the Task at Hand and Making Information Relevant to It,” *Intelligent Systems and Applications*, pp. 1-8, 2000
7. J. Beaudry, “Boundary Objects and Social Design,” Jeremy Beaudry - Social Designer & Educator, [Online]. Available: http://www.boxwith.com/2015/01/21/boundary-objects-and-social-design/. [Accessed 24 April 2016]
8. D. Snowden, “The Cynefin Framework,” Cognitive Edge, [Online]. Available: https://www.youtube.com/watch?v=N7oz366X0-8. [Accessed 24 April 2016].
9. J. Kim, “How Complexity Domain Impacts Software Development Process,” Quaterview, 9 March 2014. [Online]. Available: http://quarterview.com/?p=1091. [Accessed 24 April 2016].
10. E. J. Christie, D. D. Jensen, R. T. Buckley, D. A. Menefee, K. K. Ziegler, K. L. Wood and R. H. Crawford, “Prototyping Strategies: Literature Review and Identification of Critical Variables,” American Society for Engineering Education, 2012
11. S. Koch, A. Sheeren and N. Staggers, “Using Personas and Prototypes to Define Nurses’ Requirements for a Novel Patient Monitoring Display,” *Connecting Health and Humans*, pp. 69-73, 2009
12. S. Kvale, Interview - En introduktion til det kvalitative forskningsinterview, København: Hans Reitzels Forlag, 1997.

How can a user interface support understanding of rich data? Evaluation of an application for visualisation of arterial blood gas.

Adam Viggo Glistrup
Aalborg University
Department of Computer Science
Aalborg, Denmark
adam@adamglistrup.com

Dennis Dalgaard
Aalborg University
Department of Computer Science
Aalborg, Denmark
ddalgaard@gmail.com

ABSTRACT
Visualising rich data for highly specialised domain experts is difficult due to the complexity of the interpretation. This paper investigates how a user interface can support highly specialized domain experts' understanding by visualising rich data. We developed an application that visualises arterial blood gas which was evaluated with eight physicians at Intensive Care Unit at Aalborg University hospital. The evaluation was conducted as a comparison to their current practice through field experiments. The findings revealed that sequential data results on a graph benefits the ability to discover sudden fluctuation and understand trends over time. Information presented in a compact context increased interpretation speed and possibly quality. Visualisation must include contextual reference such as an reference interval or treatment information and display recent results. The application must be customisable to the specific domain expert and allow time to convert from current practice.

Author Keywords
Visualisation; user interface; rich data; specialised system.

INTRODUCTION
Data visualisation is a very well documented research field with many detailed areas and specialised methods for different situations. The purpose of data visualisation is to provide an approach that can aid the exploration and understanding of data. Recent developments in computing power have benefited data visualisation greatly, in terms of drawing precise models, creating complex displays, and printing in remarkable quality and high resolution [1]. However, selecting and using the correct method can be difficult. In order to select the most fitting method it is crucial to have a well-defined understanding of the content, context and the construction of the visualisation.

When the main goal is to visualise data through an IT system, the user interface of the IT system becomes a very essential part. The user interface is a link between the user and the computer which users see and interact with [2].

Data that contains a large amount of information is for the purpose of this study considered rich data. Rich data introduces an increased complexity as the interpretation becomes obscure and domain specific. Therefore, the interpretation requires domain experts to understand the data properly. This means that visualisation of rich data must support these domain experts' understanding as they are the interpreters.

The aim of the paper is to investigate how a user interface can support highly specialised domain experts' understanding by visualisation rich data. This is done through a project in collaboration with Aalborg University hospital with the goal of evaluating an application that can assist physicians by visualising large amounts of patient data in a comprehensive manner. The application that visualises these patient data will serve as foundation for this paper and evaluated. The patient data used are arterial blood gas samples collected at an Intensive Care Unit (ICU).

Following section will present paper's relation to arterial blood gas visualisation, how to interpret it and what types of visualisation exists. We then present the ICU's current practice for handling arterial blood gas results, followed by a presentation of the developed application for visualising these results. Then we describe the method for evaluating the application. The findings include a description of what we discovered during the evaluation. Furthermore, the discussion puts the findings in perspective and compare the results with other papers. Lastly, the paper ends with a conclusion of the paper.

RELATED WORK
The use of arterial blood gas results as a diagnostic tool are regularly utilised in intensive and critical patient care. There are many different methods for interpretation of arterial blood gas results and each method relies on different amounts of data derived from an arterial blood gas samples. One method suggests a very systematic approach [3], in which the results are accessed in six sequential phases. Another method proposes a stepwise approach [4] where

each step offers multiple options, creating a decision tree for interpretation of arterial blood gas results.

Visualising the results from arterial blood gas samples is not unique. However, due to the large variety in how arterial blood gas results are analysed and interpreted, there does not exists a common practice. In 2011 Alexa K. Doig and her colleagues studies one approach [5]. They created a software visualisation that visualised the three most commonly used components in an arterial blood gas analysis: pH, PaCO2 and HCO3-. In the study they evaluate their software against traditional numerical display. They found that the response accuracy when identifying acid-bases states increased significantly as well as reducing workload when interpreting arterial blood gas results [5].

Unfortunately, neither of these share similar content in terms of data structure of the result data. This make replication of these methods unsuitable and different means of visualisation are required.

Another study that works with improving ICU workflow is Kochs et. al. article [6] on information displays for improving situation awareness. This study combines information from several traditional display into a single paper prototype containing an overview of the most relevant information. The prototype was evaluated with experienced nurses and through a series of counterbalanced tasks measures the perception, comprehension, projection and completion time in order to determine the nurses' situation awareness compared to traditional methods. The study found that such integrated ICU information displays increased the nurses' situation awareness, decreased task completion time, and potentially reduce errors [6].

As for this paper, the purpose is to evaluate our application that expands on the ideas of visualising arterial blood gas results. The application facilitates both systematic and stepwise approaches for interpretation of the samples. The evaluation of the application is similar to Koch's et. al. [6], in that we also evaluate a prototype vs the traditional method through counterbalanced tasks.

OBSERVATION FORM

To better understanding their current practice, the following section will present the observation form that is their current approach for interpreting arterial blood gas results.

The observation form consists of physical papers that includes arterial blood gas samples, respiratory settings and other miscellaneous treatment information. This is their current practice of handling the information. There may be multiple observation forms associated with a single patient, depending on amount of time the patient is hospitalised. Figure 1 shows the inside of a single observation form. The papers taped together are the arterial blood gas samples and above them is the respiratory settings.

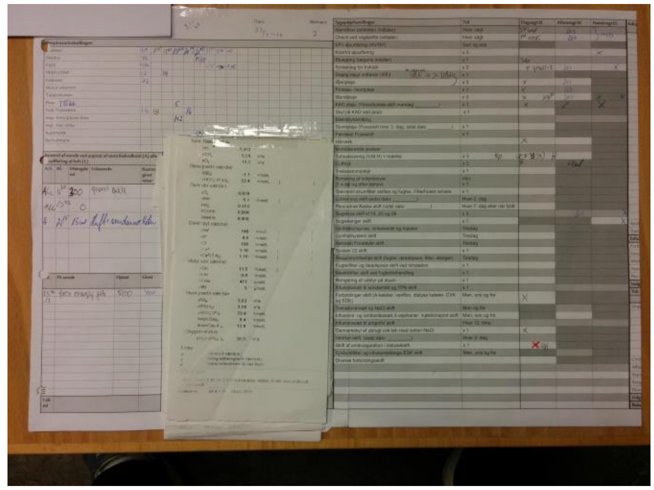

Figure 1. Picture of observation form.

Each of the taped paper represents a single arterial blood gas sample. Figure 2 depicts many papers of arterial blood gas samples bundled together as a large stack.

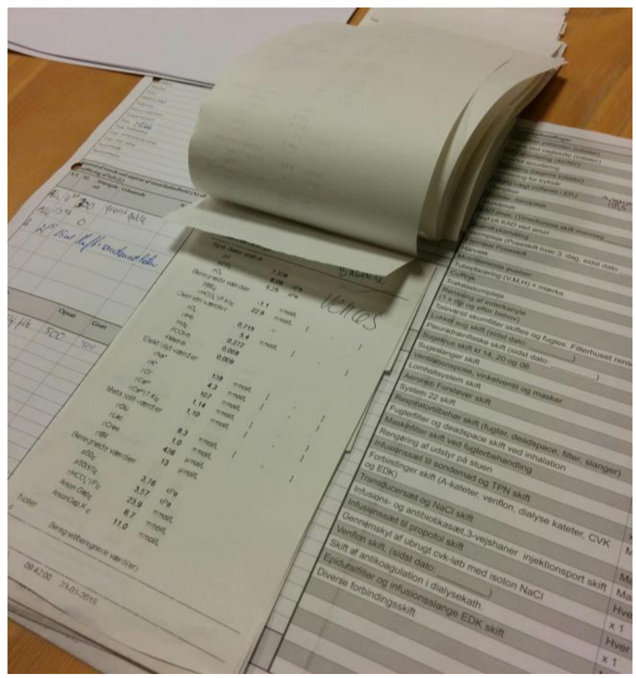

Figure 2. Picture showing amount of arterial blood gas samples included in each observation form.

APPLICATION

This section presents the application developed as our suggestion for how to improve arterial blood gas visualisation.

The evaluated application was developed in collaboration with the ICU at Aalborg University hospital with the goal of improving the practice within their department. It consists of a server, accessible through a browser, that collects arterial blood gas results from a database. The user searches for the patient via Central Person Registry (CPR) which then are processed, ordered and displayed as a list of components with corresponding graphs (see Figure 3). This includes that same information that is available from the observations form's paper strips. Each graph includes all results within a certain time period and the components' reference interval – the interval which are considered normal for the average citizen. The white area depicts the normal reference and the grey area depicts too high value or too low value. Each sample is highlighted with a minor dot that can be hovered to inspect exact details about that sample.

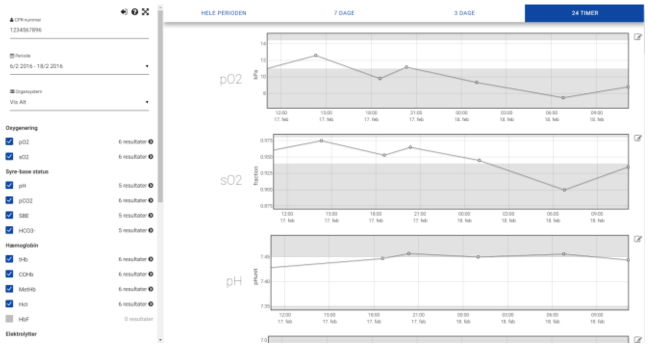

Figure 3. Initial overview when accessing a patient's results.

The user can manually, or with prefixed groups, change between which components are shown (see Figure 4) and switch between time intervals. Doing this allows the user to see specific components of the user's choice over variable period of time. It serves as a shortcut to faster gain overview of the relevant components.

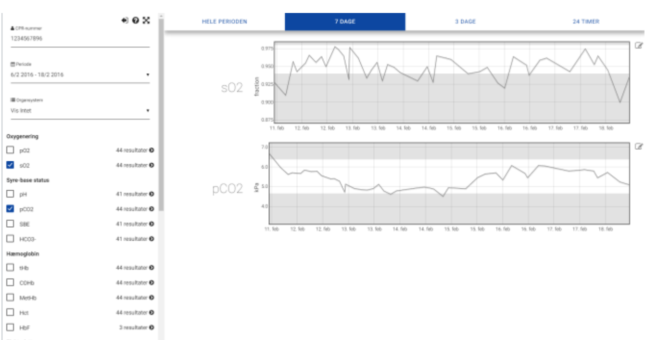

Figure 4. Components shown when selecting a specific organ system.

Additionally, the user can edit the reference interval to fit the specific patient (see Figure 5). By moving the slider on the right side, the reference interval adjusts the graph correspondingly. This feature was requested to match their current practice where they manually define the patients' reference interval each morning.

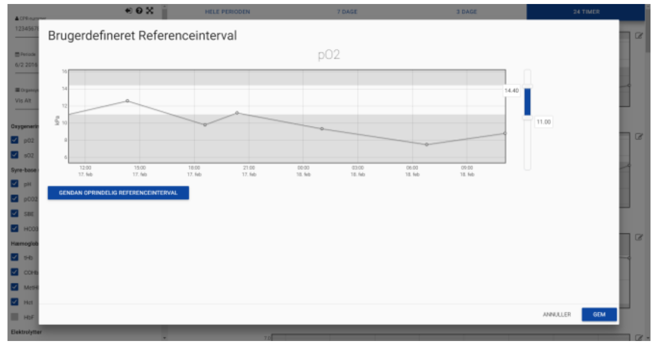

Figure 5. Option for editing reference interval.

METHOD

To investigate the problem we used the developed application to evaluate how to support rich patient data. This section describes the method used for the evaluation.

Method Development

As we have very limited medical knowledge, designing a relevant and meaningful evaluation proved to be substantially difficult. To combat this we enlisted the help of the physicians involved in the project. With their help we were able to obtain two real medical patient histories and create relevant tasks for these patients. These two patient was hospitalised for respectively chronic kidney insufficiency and chronic pancreatitis. This caused a new challenge as the medical oriented outcome would be impractical for us to analyse. Therefore, the physicians participating in the evaluation was asked to gauge their own answers. This measure of self-evaluation along with qualitative statements served as the actual results of the evaluations. This means the medical correctness of specific answers are not taken into account. Prior to the actual evaluation, we conducted a pilot test to ensure the approach made sense for physicians.

Materials & Settings

Materials used in the evaluation was a laptop with the application and patient data. The difference between their personal computer and our laptop would be very insignificant but our laptop provides consistency as it prevents uncertainties such as bugs or visual differences. The patient data was fetched from a production database and cached for consistency of the evaluation. We used a laptop and paper notes to help ourselves follow the procedures. The evaluations took place at the participant's individual office and varying meeting rooms at Aalborg Hospital during normal work hours. This is the intended location for the application's usage and increases the realism of the evaluation.

Participants

Three persons were be present at each evaluation: A physician, a facilitator, and a note taker. The participants included eight different physicians which were working at ICU at Aalborg University hospital. The physicians were selected by ICU's Chief Physician and PhD student and represents common users for the application.

The facilitator ensured procedures are maintained and the physician followed the scope of the evaluation. The note taker had responsibility for the technology was working and took notes during the evaluation. For the sake of consistency we used the same facilitator and note taking across all evaluations.

Procedure

The evaluations were initialised with an introduction of the project and aim for the evaluation. It established a non-formal setting between the participant and facilitator and ensured the participant was comfortable in the situation.

After this, the participant was asked to solve a task based on one of the cases with either the application or the observation form. The participant was given five minutes to takes notes about the patient's medical development during the hospitalization. The participant was asked to think-out-loud while performing the task. After the time expired, the participant was asked to solve the other case using the application or the observation form, oppositely of previous case. When using the application, a brief introduction of the application was presented to the participant. This includes a demonstration of all relevant features and the participant using the application, until the participant felt comfortable with it. This was to counter the lack of experience the participant had with the application which may result in a disadvantage. To counter the aspect of the participant learning during the tasks, we will rotate the sequence of the cases and the application/observation form for the eight participants. The rotation uses following structure:

No. participant	First task	Second task
Participant 1 & 5	Case A (Observation Form)	Case B (Application)
Participant 2 & 6	Case A (Application)	Case B (Observation Form)
Participant 3 & 7	Case B (Observation Form)	Case A (Application)
Participant 4 & 8	Case B (Application)	Case A (Observation Form)

Table 1. Rotational structure of case and display method.

After both tasks, the participant was asked to fill a questionnaire answering the following three questions on a scale from 1 to 5 for both cases:

1. How confident are you in your answer?
2. How comprehensive do you think your answer is?
3. How much influence did the time limit affect your answer?

The questionnaire was followed by a semistructured interview to gain insight in the reasoning for their answers and the participant's qualitative opinion on the application. Finally, the participant was offered to use the application without any related task to freely comment on their perspective and opinions on the matter.

Data Collection

Notes were taken by the note taker and sound was recorded as well. Sound recordings were used throughout the entire evaluation to ensure no data was forgotten. The questionnaires were collected for later analysis.

Data Analysis

The goal of the analysis was to interpret the participant's confidence in using the application and the time required to gain proper comprehensive understanding of the patient through the application. This was analysed in relation to the observation form to gain a relative understanding of these goal.

The questionnaires' answers was compared within the same participant within the same question between the two cases (see Figure 6). The analysis will interpret the difference between these answers assisted by a meaning condensation [7] conducted from each interview. Additionally, the analysis includes interpretation from the open interview to gain better quality of the reasoning.

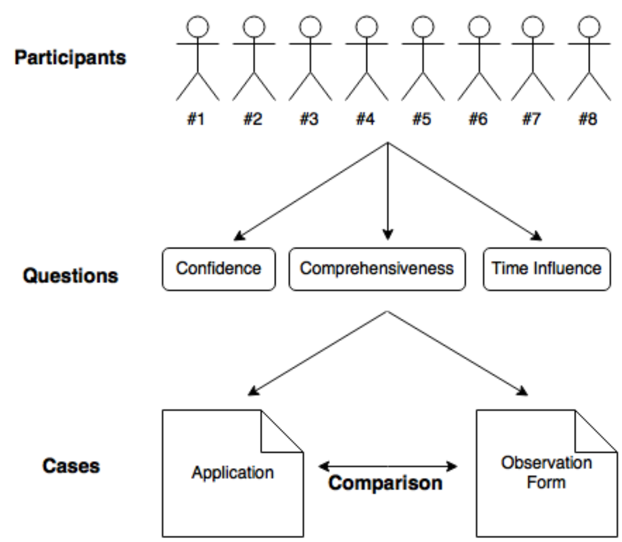

Figure 6. Comparison of the questionnaire.

RESULTS

The results are structure through four subsections: questionnaire, improvements, issues and observations. The questionnaire subsection will present the result of the three questions. The three other subsections are our findings based on entirety of the all the evaluations. Improvements includes arguments for how to improve the application, both in relation to the observation form and isolated for further development. Issues includes aspects of the application that is found to require more attention and must be considered for optimisation. Observations are findings that are neither positive or negative remarks but relevant for the visualisation.

Questionnaires

The results from the questionnaires are listed below. Each figure shows one question asked twice by each participant, once for each display method (application or observation form). The figure shows the difference in responses according to each display method. Only the difference is relevant as the absolute value has no common meaning between the participants. Therefore, the difference represents which of the methods the participant considers better in relation to question. All figures (see Figure 7, Figure 8, and Figure 9) are illustrated with positive numbers as preference towards the application application and negative number as preference towards the observation form. The following shows three examples of the results for the question about confidence:

1. Participant 1 answered 5 to both the application and the observation form. This result in 0 (no difference)
2. Participant 4 answered 5 to the application and 4 to the observation form. Therefore, there is a single point difference in favor of the application which is 1.
3. Participant 6 answered 3 to the application and 4 to the observation. Therefore, there is a single point difference in favor of the observation form which is -1.

Question 1: Confidence

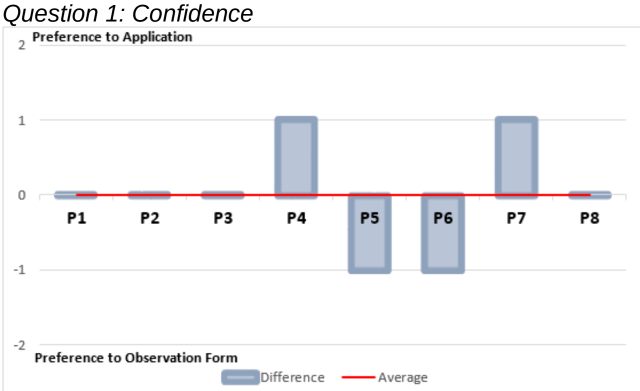

Figure 7. Confidence difference between the two methods.

Figure 7 shows four of the participants had the same confidence in their answers for both methods. The two who had more confidence in the application argued that the sequential listing of blood samples verifies their analysis and ensure a little more confidence. Participant 4 said: "*I'm more confident I haven't missed a drop somewhere*". The two other who prefer the observation form justified it with force of habit towards the observation forms . Participant 5 said he had less confidence in the application as the components are ordered differently compared to the observation form.

Question 2: Comprehensiveness

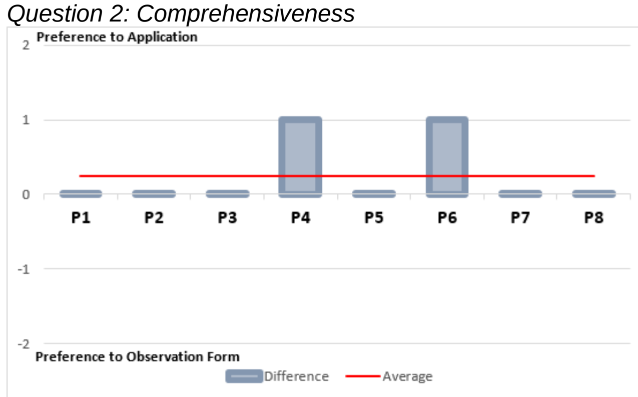

Figure 8. Comprehensiveness difference between the two methods.

Figure 8 shows that the majority of the participants think both methods yields equally comprehensive answers. The two who think the application allowed more comprehensive answers, said it was because the graphs gives an instantaneous picture of a component hence being able to answer better. Participant 6 said: "*The graphs can help to identify significant events*".

Question 3: Time Influence

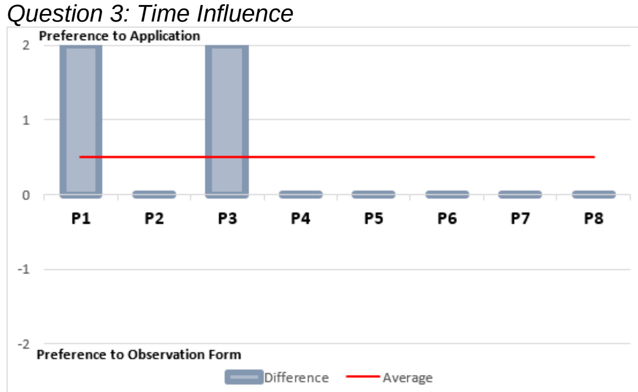

Figure 9. Time influence difference between the two methods.

Figure 9 shows that most participants did not feel the time constraint influenced their answers, whether they used the application or the observation form. However, two participants suggested the application was better. These two participant did not use the full time with the application but did with the observation form. Participant 1 said: "*It is much faster because I got the time axis instead of flipping through the pages*".

Improvements

Sequential display of components improved understanding of trend

The most convincing finding was the positive response to the graphs that displays a sequence of results for each component. All participants was able to interpret a patient's component's development within few seconds. Participant 1 said: "*It is the tendency over time that is interesting*" supported by participant 7: "*It is the tendency I'm looking at.*

I rarely work with a single value". The graphs helps the immediate understanding a component's development as participant 2 said: "*The graphs gives a picture whether it's going the right or the wrong way*". Participant 8 used this to quickly identify a significant event: "*Something here stands out which is his low pO2 at 4 o'clock*".

"*It is obvious; we need a trend function*" - Participant 7

The observation form was time consuming

One of the reasons for the overwhelming positive feedback of the graphs can be found in the comparison their current practice "*It is a good layout, compared to flipping through fifty pages*" - Participant 4. We observed all of the participants spend a lot of time flipping through the pages in the observation form. For instance did it take nine seconds for participant 1 to conclude a patient has increasing oxygen between two samples. The participants often attempted to gain a mental visualisation of the paper by remembering the values on different pages. For instance did participant 5 and 7 flip through all the pages multiple times as it was difficult to maintain a mental model of all the different components at the same time. Another observation was that several of the participants did not attempt to solve the task they were given with the observation form but rather explained how they would do it. This indicates that the actual process of comprehending the data was too time consuming to perform.

"*It (observation form) takes more time because you have to flip through the pages*" - Participant 3

New conclusions were discovered through the application

According the the questionnaires, only two participants believed the application provides more comprehensive answers. However, we observed multiple instances of the participants reaching new conclusions when using the application. Participant 4 commented within a few seconds of looking at a specific component: "*I can wonder why they haven't given him more oxygen here when he's running low*". It seems like the visualisation reveals a development in the component that may have required a different medical action. Participants 8 reflected upon the large amount of time between two samples: "*I think it's noticeable they wait for so long to take a new sample (...) I'd expect to take another sample within a half hour because it goes very downhill for him at this point*". Additionally, there were made conclusions on multiple occasions that the sample was plain wrong: "*There is a faulty source for creatinine because it cannot vary this much*" - Participant 3. This was a well-known issue they had as some components are calculated from manual inputs. The application allowed to quickly identify this.

"*He's falling in hemoglobin. I believe he should have been given some blood here*" - Participant 3

Reference interval helped set the context for the results

We observed that the reference interval provided a good context for the results. As all the participants were very familiar with the reference interval of all components, the participants used the visual representation much more than the actual numbers. Participant 3 was able to assess all samples of a given component within few seconds by relating the samples' position to the grey reference interval area. Participant 7 also found it beneficial: "*It is good you have the white area (normal reference interval)*". Participant 6 commented on the option to customise the reference interval: "*Fine that you can individualise the reference interval for this patient in regards to coordination with the nurses*".

"*It is a fine representation and the curve is good with the reference interval*" - Participant 1

Useful to gain overview of new patients

Some participants expressed satisfaction with the application in the context of gaining information about a new patient. Participant 5 said it provides a faster overview when conducting hospital rounds, which occurs in the beginning of a physician's shift. Participant 8 also said it provides a good overview that enables insight into what treatments has previously worked for the particular patient.

"*When we conduct hospital round and must acquire an overview of what has happened the previous day, you could definitely used this (application) to look back at the development*" - Participant 8

Issues

Different components requires different visualisation

The application visualises every component the same way but the components' medical usage are more complex than implemented and some may require different individual visualisation. For instance, the component creatinine only makes sense to visualise over a long period of time as it does not change much over a day. Oppositely, components related to respiration only matter within a very short timeframe as they do vary a lot. This was commented on by multiple participants when they swapped between the 24 hours view and the 7 days view. Some components may not require a graphical visualisation as participant 8 suggested: "*I don't necessarily want everything presented as graphs*".

"*Most of them (components) are relevant within 24 hours but the hematocrit and the creatinine would be relevant to see over a longer period*" - Participant 1

Missing treatment information

The evaluation revealed a lack of medical context in the application. All participants requested information about respiratory settings, which is how much oxygen the patient is given: "*Maybe it could be interesting to see what respiratory settings the patient is on, on the time axis*" -

Participant 1. “*What I need in the application is the respiratory settings*” - Participant 3. These settings was part of the observation form and are relevant to the components displayed in the application. Participant 6 elaborated on this request further by asking for more treatment information about what has been given to the patient and at which time. The participant gave a few examples: How much oxygen the patient is receiving, whether the patient uses an oxygen mask and how much blood the patient has received.

“*I usually also look at the respiration settings*” - Participant 2

Lacking immediate understanding of recent samples

Some of the participants discussed the issue of not being able to get an immediate display of the most recent samples. In comparison to the observation form, they have all the recent results collected on a single piece of paper. Participant 1 suggested it could appear next to the graphs: “*Maybe you could have the actual value next to it (the graph)*”. The most recent samples are considered especially relevant as they are the closest information they can get about a patient’s current situation.

“*It can be difficult to see the actual value, which is what we often use arterial blood gas for*” - Participant 2

Difficult to view all components

The user interface only displays about three components at the same time although there are about 25 components for every patient. This makes it difficult to get an overview of all components. For instance, did it take a while for participant 2 to identify where lactate is located. Participant 4 scrolled through all of the components to get an overview which was time consuming, “*I need to scroll back and forth to get all the components at the same time*” - Participant 8. Participant 7 tried to solve the issue by initially disabling all components the participant considered irrelevant. The issue may also be related to the their current practice where they are used to viewing all components on a single page. This is supported by participant 6 who requested to see an entire list of values for all components at a given time (identical to current practice).

“You lose the overview as you can only see a few components” - Participant 2.

Components are relevant in relation to each other

We discovered that some of the components are used, almost exclusively, in relation to other components. The application does not support this very well as we had multiple requests for seeing certain components grouped together. Participant 2 requested: “*I would like to combine multiple curves, so that you can see them on the same graph*”, supported by participant 6 who also wants multiple components merged together. As an example, participant 3 suggested: “*When you hover over pO2, you should see the corresponding Fio2 value*”.

“*You should group the components that makes sense together, and they must be visible on the same screenshot*” - Participant 7

Observations

Largely different usage

We found large differences in how each participant handled the tasks they were given. Although we cannot conclude on the medical technical answers, we did observe significant differences in the use of both the application and the observation form. Participant 1 used the hover feature to inspect specific samples a lot while participant 7 did not and explicitly expressed: “*All I need is the graphs*”. For the observation form we notice it varied whether the participants started from top (most recent sample) or from the back (first sample taken that particular day).

Participant 3, 5, 6, 7 and 8 requested another order of the components as it did not fit their expectation of how relevant they were. The noticeable observation is that all requests were different, hence the participants prioritise components differently. This was confirmed as the talk-out-loud method revealed the participants looked at different components to solve the same task. Participant 3 elaborated that the same analysis can be conducted by looking at other components which he solved by manually disabling the components the participant did not find relevant. Another suggestion from participant 7 was to create user defined views that suits each physician.

Difficult to adjust to changes

The participant has used the observation form as their current medical practice and are used to this way of analysing samples. This makes it difficult to truthfully compare the methods to each other as participant 2 said: “*Both parts (application and observation form) has it’s pros and cons. It is also because we are trained in looking at the stack of papers*”. It reveals that the application is a big step away from their current practice, which naturally has some adjustments disadvantages that the participants recognised: “*If this existed (application), it would be used, although it would requirement some adaptation*” - Participant 2 and “*If I wasn’t used to working with arterial blood gas strips, this (application) would be a hundred times better*” - Participant 8. Another example is participant 5 who requested the possibility to see samples the same way it is displayed in the observation. As previously mentioned, participant 6 also wanted this feature. Whether these requests were caused by habit or actual useful suggestions is difficult to conclude.

“*I like the old way better (observation form). I don’t know why. Maybe it’s because it’s a habit*” - Participant 4

DISCUSSION

This section delves deeper into the results of our evaluation by presenting the knowledge that was derived from working on this project and discusses its relevance in relation to the related work.

The results of this study can be categorised in three categories: quantitative data from the questionnaires, qualitative feedback and observations about the application vs observation forms and qualitative feedback about specific features of the application.

Questionnaire

While the results from the questionnaire did not generate any large variations in the responses, the variation that was recorded was mainly positive towards the application. Only two participants responded negatively toward the application when asked about the confidence in their answers. However, when asked about this both participants contributed their negative response to unfamiliarity with the application. We cannot derive much from the questionnaires alone but can use them as a reference to our other findings. It can be used a baseline to argue that there are both negative and positive aspects with the application.

Application vs observation form

The evaluation revealed several strong arguments as to why the developed application has potential to aid physicians when diagnosing and monitoring patients through arterial blood gas analysis. The clearest advantage identified was the improved understanding of trends: All but one of the participants highlighted the visualisation of trends over time as a clear positive feature. Some participants even compared it other applications with similar functionality with other patient information, although, none of them were as detailed as the application. Two of the participants even reached new conclusions based on the visual presentation of the data; conclusions that did not seem to have been reached when the patient was originally treated. However, the validity of these conclusions remain unknown. Another strong point that was highlighted several times was the visualisation of reference interval. The observation form did not contain this information but the evaluation clearly revealed the usefulness of this information. The functionally to define an individual reference interval for each patient might even make it more relevant, which was confirmed by a few of the participants.

In regards to the observation form, some of the most notable discoveries was that it currently contains more information than the application provides, specifically respiratory settings and patient treatments. However, the observation forms is hopelessly outdated and time consuming to interpret. Despite the questionnaires not revealing a very significant different, the time spent solving the tasks reveals it might be more noteworthy. In the evaluations we observed that most of the participants would spend all time available to solve the task when using the observation form, but when using the application they would finish the task ahead of time. Whether this time different can be contributed to lack of respiratory settings and patient treatment information or faster understanding through the application is difficult to determine. Another observation was that several of the participants spent a significant amount of time shifting back and forth between the arterial blood gas strips. This is comparable to Sven Koch's work with ICU information displays [6], in which he concluded that: "*ICU information display increased ICU nurses' situation awareness and decreased task completion time for medication management, assessments of the patient's state and team communication*" [6]. We also found that the physicians had a very difficult time to adjust to the changes which contributes to a larger disadvantages for the application. Despite all this, the comments and findings have been rather positive towards the application. This limitation causes an unfair disadvantages that makes the comparison shifts in favor of the observation form.

Application feature feedback

The application received feature specific feedback that incentivises further discussions. A surprising discovery is the importance of respiratory settings that is not included in the application. Almost all participants used this information as an essential part of their analysis. The importance of this information is supported in Heidi Simpson's article, and she includes, efficiency and adequacy of ventilation; work of breathing and pulse oximetry [3] as some of important factors to incorporate when evaluating arterial blood gas results.

Another interesting point was the vastly different diagnostic process used among the participants. The order of the components in the application was defined by the physicians involved in the project. However, multiple of the participants commented that the order was not ideal or did not match their individual diagnostic process. Interestingly, these participants all suggested different orders. A few of the participant even requested the ability to save the customised view and rearrange the order. This indicates that there is no best practice on how to interpret arterial blood gas results. Earlier we presented the work of Heidi Simpson [3] and Barnette & Kautz [4] who respectively argue for two very different methods. However, it would seem that the process is far more individual than initially assumed, as different departments and different types of physicians have very different approaches to arterial blood gas results. This, in combination with the feedback from the evaluation, strengthens the need for customised views.

CONCLUSION

The evaluation revealed some positive and negative findings when visualising rich data in a user interface. The most prevalent benefit was the ability to discover sudden fluctuations and understand trends over time by displaying a sequence of results on a graph. Another valuable attribute is that the information is presented in a much more compact context, making the interpretation significantly faster and possibly of better quality. The reference interval served as an example of a great contextual indicator.

To improve the understanding of the data, the application would require some adjustments to include treatment information, better display of recent results, customisation specific to the physicians and time to convert from the observation forms.

The evaluations compared two radically different methods of presenting rich data, and the participants was still unfamiliar with the application despite a thorough introduction. This might have lead to a unfair comparison. However, the feedback and comments suggests that if the participants was given more time to familiarise themselves with the application and adjust their diagnostics process accordingly, then the advantages of the application would be even more significant.

ACKNOWLEDGEMENT

We would like to thank Aalborg University hospital for their involvement in the process. Specifically, we would like to thank the participants of Intensive Care Unit and the physicians who helped establish evaluation. We would also thank Jan Stage for providing supervision throughout the entire project.

REFERENCES

1. A. Urwin, “Good Graphics?,” in *Handbook of Data Visualization,* Berlin, Springer, 2008, pp. 57-78.
2. D. Benyon, Designing interactive systems: a comprehensive guide to HCI and interaction design, Harlow: Addison Wesley, 2010.
3. H. Simpson, “Interpretation of arterial blood gases: a clinical guide for nurses,” *British Journal of Nursing,* vol. 13, no. 9, pp. 522-528, 2004
4. L. Barnette and D. D. Kautz, “Creative Ways to Teach Arterial Blood Gas Interpretation,” Lippincott, Williams & Wilkins, 2013.
5. A. K. Doig, R. W. Albert, N. D. Syroid, S. Moon and J. A. Agutter, “Graphical Arterial Blood Gas Visualization Tool Supports Rapid and Accurate Data Interpretation,” *Computers, Informatics, Nursing,* vol. 29, no. 4, pp. 53-50, 2011.
6. S. H. Koch, C. Weir, D. Westenskow, M. Gondan, J. Agutter, M. Haar, D. Liu, M. Görges and N. Staggers, “Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: A prospective randomized controlled study,” *international journal of medical informatics,* no. 82, pp. 665-675, 2013
7. S. Kvale, Interview - En introduktion til det kvalitative forskningsinterview, København: Hans Reitzels Forlag, 1997