Understanding system features through information visualisations to improve product development: An exploratory study with data analysts

Agzam Idrissov[1], Stanko Škec[2], Anja M. Maier[3]

[1] *Engineering Systems Design, DTU - Technical University of Denmark*
agzid@dtu.dk
[2] *Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb*
stanko.skec@fsb.hr
[3] *Engineering Systems Design, DTU - Technical University of Denmark*
amai@dtu.dk

Abstract

Information visualisations, such as diagrams and charts, support analysis of large amounts of information. As product development is increasingly becoming digitalised and while an abundance of information presents challenges, there are also increasing opportunities for engineers to become even more data-savvy and to leverage the power of information visualisations. While there exist various information visualisations for decision support in product development, these representations are often being created in an ad-hoc manner and often specifically for each task. This becomes particularly challenging in systems-oriented design where multiple domains and views need integration. In this paper, we examine whether certain types of information visualisations facilitate understanding of characteristic features of systems, such as hierarchy, relationships, patterns and processes. Results from our study with data analysts suggest that information visualisations are recognised as representing the aforementioned characteristic system features, even by study participants without prior familiarity with the visualisations used, i.e. network diagram, treemap, Sankey diagram, co-occurrence matrix, and treemap.

Keywords: visualisation, design process, product development, product architecture, system architecture

1 Introduction

Large amounts of product information are being collected throughout the product lifecycle and even more so now, given that product development as a whole is becoming increasingly digitalised. In addition to the sheer volume of data, product information is heterogeneous, consisting, for example, of unstructured texts of customer needs, tables of requirements, or numerical data about physical characteristics of a product, and it is also constantly evolving. Navigating in this new world of data and information, it is not uncommon for engineers of today

to get involved with its handling and analysis. Therefore, making sense and communication of large amounts of heterogeneous product information in product development needs to be supported.

Information visualisations, defined as "*computer-supported, interactive, visual representations of data*" (Card, Mackinlay, & Shneiderman, 1999), are one of the ways to facilitate understanding of complex product information. Due to their benefits for individual and shared sense-making (Card et al., 1999; Larkin et al., 1987; Tufte, 2001; Vessey, 1991), interactive, data-driven information visualisations (henceforth referred to as „information visualisations") are being widely embedded within the new generation of Product Lifecycle Management (PLM) tools (Allanic et al., 2014, Krause, 2007). Examples of information visualisations include network diagrams, treemaps (Shneiderman, 1992), co-occurrence matrices (Steward, 1981), Sankey diagrams (Riehmann et al., 2005), parallel coordinates plots (Inselberg, 1985) and other visualisations (Heer et al., 2010).

Despite increasing adoption of information visualisations in practice, they are often prepared in an ad-hoc manner, i.e. for each design task afresh, which often results in duplicate efforts when creating visualisations and suboptimal choices of visualisation techniques to represent the information at hand. These issues can be addressed if the fundamental properties of visualisations are known, and this knowledge is deliberately and systematically used within PLM systems and design support tools.

One way to examine how fundamental properties of visualisations match the represented product information is by treating products and related information as systems (Crawley et al., 2015). By understanding characteristic features of systems and mapping them to visualisations that can represent these features, visualisations can be prepared more appropriately. In prior work, we elicit characteristic features of systems as relationships, hierarchies, patterns and processes, and then map them to existing information visualisations (Idrissov et al., 2020). For example, tree diagrams are more appropriate than network diagrams to show hierarchy (Novick and Hurley, 2001). Similarly, a matrix visualisation is shown to be efficient at representing a large number of relationships between elements (Keller et al., 2006). Thus, a mapping between characteristic system features and information visualisations that are suitable to represent them permits a more informed selection of appropriate visualisations to support everyday product development tasks.

Thus, in this paper, as an extension of the said prior work (Idrissov et al., 2020), we conduct an exploratory empirical study with expert data analysts to understand how information visualisations could enable an understanding of characteristic system features. This understanding could then be used more systematically during product development. The remainder of the paper is organised as follows. First, a brief literature background on information visualisations in product development are presented. Second, the study method for the empirical study investigating whether certain information visualisations afford recognition of certain system features is reported. Third, the findings of a questionnaire study with data expert analysts investigating accuracy, familiarity, and use of specific information visualisations representing characteristic system features are examined. Finally, the paper concludes by discussing the findings and implications for research and practice.

2 Background

2.1 Information visualisations in product development

Product information can consist of a large number of interconnected elements, non-trivial to analyse (Summers and Shah, 2010). First, product information may be heterogeneous with descriptions including information on the goal (e.g. customer needs, requirements, functional decomposition), process (e.g. tasks, activities) and product (e.g. physical or software product architecture) (Danilovic and Browning, 2007). Second, such product information can be represented in different levels of granularity (Sim and Duffy, 2003). Often, hierarchical decomposition is used to help reason about the often large systems (Alfaris et al., 2010; Hehenberger, 2014). Finally, product development is a highly dynamic process, continuously changing product information at hand. For instance, a change in a single requirement can result in redesigning significant parts of an already developed product architecture (Clarkson et al., 2004).

One way to deal with such complexity of information is through the use of information visualisations. Research suggests that information visualisations facilitate common visual cognitive tasks, such as exploring outliers and relationships between variables (Inselberg, 1985), providing a fast overview of data (Lanzenberger et al., 2005), of facilitating pathfinding and pattern recognition tasks (Pohl et al., 2009). Such effects are further magnified by two fundamental properties of information visualisations: rearrangement and interactivity (Spence, 2001). While rearrangement of visualisation elements enables multiple views on the same information, interactivity provides users with control over the arrangement of the visual elements. For instance, sorting and filtering elements in a list or zooming into an area on a map are examples of when both rearrangement and interactivity are applied.

A brief overview of engineering design literature shows that a variety of information visualisations support the product development process. Eppler and Kernbach (2016) propose the use of information visualisations as boundary objects (Star and Griesemer, 1989) to improve communication between design participants. The Design Structure Matrix (DSM) (Eppinger and Browning, 2012) is often used, e.g. to predict change propagation (Clarkson et al., 2004). Treemap diagrams help to visualise a hierarchy of design alternatives (Yan et al., 2012) and the Design Rationale Editor tool (Bracewell et al., 2009) is based on a tree diagram and visualises various concept alternatives. A more detailed overview of information visualisations used in product development is presented in Idrissov et al., (2020).

2.2 Information visualisations for characterising system features

Despite the use of a variety of information visualisations in product development, their creation remains an ad-hoc process, tailored to specific tasks. Yet, the number of potential tasks in engineering design is large (Sim and Duffy, 2003), which makes creating appropriate visualisations an often non-reusable, time-consuming process that requires searching for visualisation layouts and interactive techniques that would appropriately convey the product information at hand. If the constructed visualisation does not match the product information, it may have suboptimal representation efficiency and cause unnecessary cognitive load on the users (Anderson et al., 2011).

To address this problem, product information can be understood and represented as a system, "*a set of entities and their relationships, whose functionality is greater than the sum of the individual entities*" (Crawley, 2015). Following such a view, in prior work (Idrissov et al.,

2020), characteristic features of systems, namely **relationships (interactions) between system entities, hierarchies, patterns and processes** are mapped on to information visualisations that can appropriately represent said features. For instance, in the case of a cleaning robot, system entities can be functions (e.g. "to move forward" or "to suck air") or parts (e.g. motor, brush, casing, wheels), which *interact* with each other in numerous ways – the brush causes vibration of the casing, the motor transmits power to the wheels, etc. Besides, these entities might be *hierarchically* decomposed to represent lower levels of granularity – the function "to clean floor" can be broken down into lower-level functions "to swipe dust" and "to suck air". Further, information about the system's performance can exhibit certain *patterns* that help to discover insights about the designed product. Whereas relationships, hierarchies and patterns describe the structure of the system, *process* representations are critical for understanding the dynamic behaviour of the system. Process representations, such as Sankey diagrams or flow-charts, show changes in a system – e.g. the flow of energy from one element to another or the sequence of activities.

While in said prior work, the foundation of the mapping between characteristic system features and information visualisations were laid out (Idrissov et al., 2020), an empirical evaluation of the proposed mapping is required to understand to what extent the selected information visualisations facilitate understanding of characteristic system features. Thus, in this paper, we aim to examine whether the information visualisations selected assist analysts in understanding characteristic systems' features. A questionnaire study was conducted with 21 data analysts. The following section describes the study methods.

3   Methods

To investigate whether information visualisations represent and facilitate recognition of characteristic system features, we conducted a questionnaire study involving 21 expert data analysts, participants of the first H2020 EURITO workshop ("EURITO: EU Relevant, Inclusive, Timely, Trusted, and Open Research Innovation Indicators", 2018). In the sample, the average work experience of the participants is approximately 18 years, and the majority of participants (13 out of 21) hold an economics degree. The study participants did not receive monetary compensation.

The main research question reads:
Q1. In what way do information visualisations represent and facilitate recognition of characteristic features of systems?

Accordingly, the first sub-question was formulated as follows:
Q1.1. Can participants correctly identify characteristic system features represented by information visualisations?

Then, to account for cases where participants were previously familiar with the visualisations, which could make them more likely to correctly match visualisations to characteristic system features, the second question was posed.
Q1.2. Are participants' abilities to correctly identify characteristic systems' features correlated with their previous familiarity with the visualisation?

Building on earlier work summarised in Idrissov et al. (2020), were the four features characterising systems were elicited from literature, for the study reported here, five visualisations that together represent all four characteristic systems features were selected

(Table 1). These visualisations were chosen on the basis of their frequent appearance in engineering design literature.

Table 1. Mapping information visualisations and characteristic systems' features

Systems' Features	Information Visualisations
Relationships, Hierarchy, Processes	Network diagram
Hierarchy, Patterns	Treemap (e.g. Shneiderman, 1992)
Processes, Relationships	Sankey diagram (e.g. Riehmann et al., 2005)
Relationships, Patterns	Co-occurrence matrix (e.g.Steward, 1981)
Relationships, Hierarchy	Tree diagram

In the questionnaire, participants were asked to match the provided versions of the visualisations with the characteristic system features they thought these visualisations represent.

This question was formulated as „*Please choose a visualisation that is the most appropriate for the following tasks*" with characteristic system features as possible answers. The visualisations had a size of 2x2 cm and black and white colours and were positioned on top of the respective questions. No definitions of characteristic system features were given to the participants prior to the study, except for the label – relationships, hierarchy, processes, patterns – as listed in Table 1. The study participants could assign several characteristic system features to each visualisation.

Printed and on paper, study participants were presented with the five information visualisation types in abstract form, i.e. without showing any data values. Figure 1 shows the visualisation examples included in the questionnaire: network diagram, treemap, Sankey diagram, co-occurrence matrix, and tree diagram.

Network diagram (graph)   Treemap   Sankey diagram   Co-occurrence matrix   Tree diagram
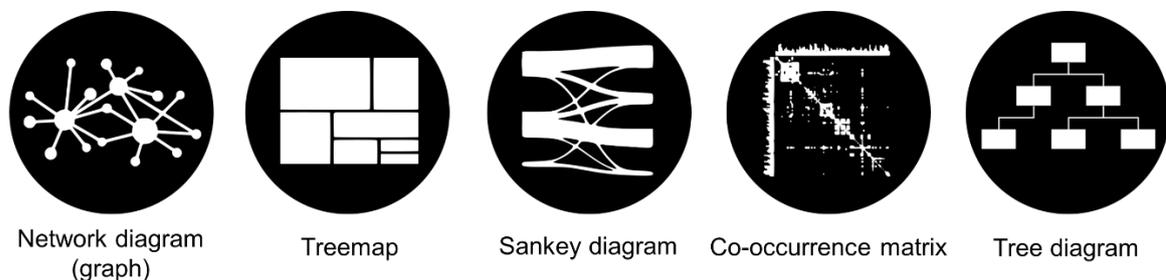
Figure 1. Information visualisation types used in the questionnaire

To investigate the effects of familiarity with the specific visualisations in question, we asked the participants to indicate whether they have previously used these visualisations and asked: „*How often do you encounter these visualisations in your work-related documents?*". Similarly, we asked whether participants use these visualisations in their work: „*How often do*

you actively use these types of visualisations to support your decision-making tasks in your work?". Answers to both questions were given on a 5-point Likert scale, where a minimum of one point was given to „never", and the maximum of five points was given for „always". The average familiarity with visualisations was calculated by dividing the sum of Likert scale points of the answers for that visualisation by the total number of responses.

Having used the mapping between characteristic system features and visualisation examples in Table 1 as the ground truth, we measured the accuracy of each user's ability to match the visualisations to the respective characteristic system features, taking into account incorrectly identified and incorrectly rejected answers (Fawcett, 2006):

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

where TP is a rate of true positives (correctly identified), TN – true negatives (correctly rejected), FP – false positives (incorrectly identified), FN – false negative (incorrectly rejected).

Finally, Pearson correlation coefficients were calculated first between accuracy and familiarity and then between accuracy and work-related usage for each visualisation type. The Pearson coefficient determines the strength of relationships between two dimensions of data and can result in a value between -1 and 1, where -1 indicates a strong negative correlation, 0 indicates no correlation and 1 indicates a strong positive correlation.

4 Results

Questionnaire answers revealed that study participants demonstrated high average accuracy (72.1%) when matching system features with appropriate visualisations, as illustrated in Figure 2. Two participants reached 95% accuracy. As shown in Figure 3, on average, participants achieved the highest accuracy with the network- and Sankey diagrams (>77%), while performing poorer when determining system features represented by a co-occurrence matrix (60.7% accuracy). This relatively poor accuracy when dealing with co-occurrence matrices could be a result of media constraints – as the questionnaire was conducted using a paper format, determining system features of visually dense co-occurrence matrices shown by a 2x2 cm black and white icon may have been challenging for the participants.
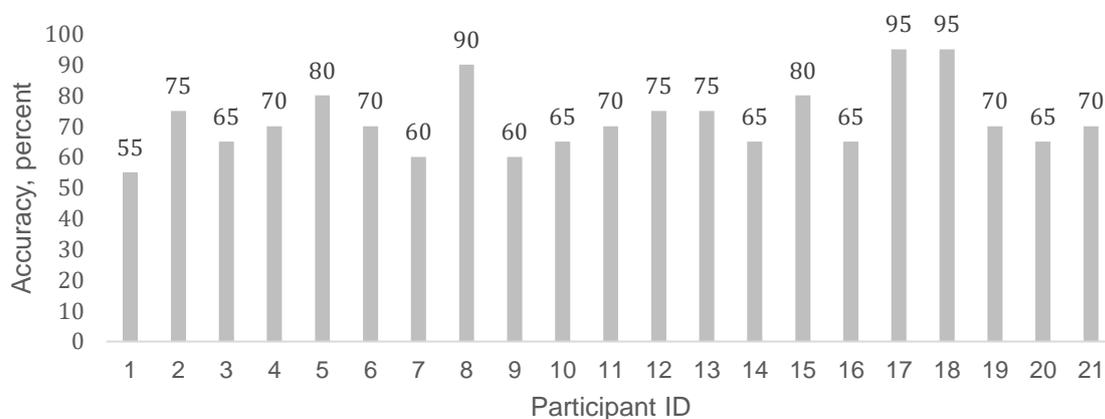
Figure 2. Accuracy per participant
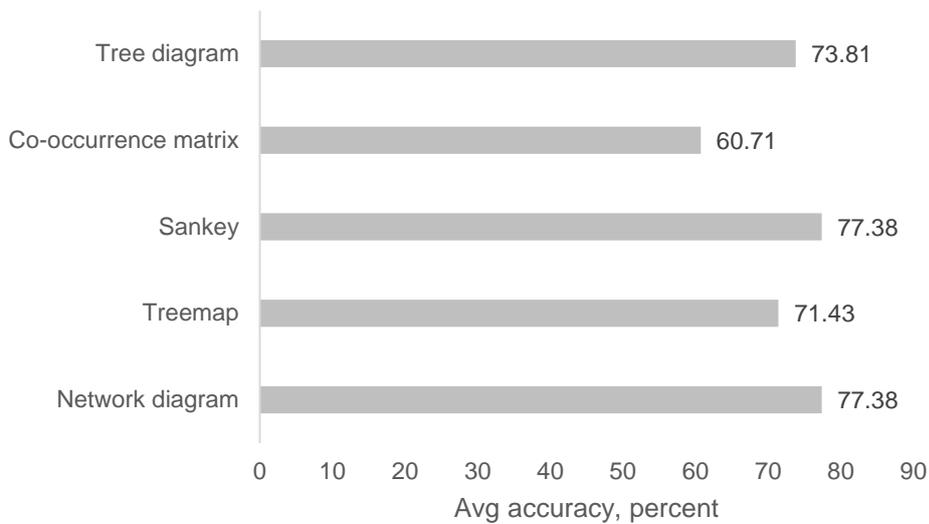
Figure 3. Average accuracy per visualisation type

Figure 4 illustrates that study participants were most familiar with the network- and tree diagrams (average familiarity > 3.0 point in Likert scale), while the Sankey diagram was the visualisation participants were least familiar with (average familiarity ≈ 2.3 in Likert scale). Seen in Figure 5, work-related usage of visualisations displayed similar results: while network- and tree diagrams were most used (average work-related usage > 2.4 in Likert scale), the Sankey diagram was the least used visualisation among all (average work-related usage ≈ 1.6, measured on a 5-point Likert scale).
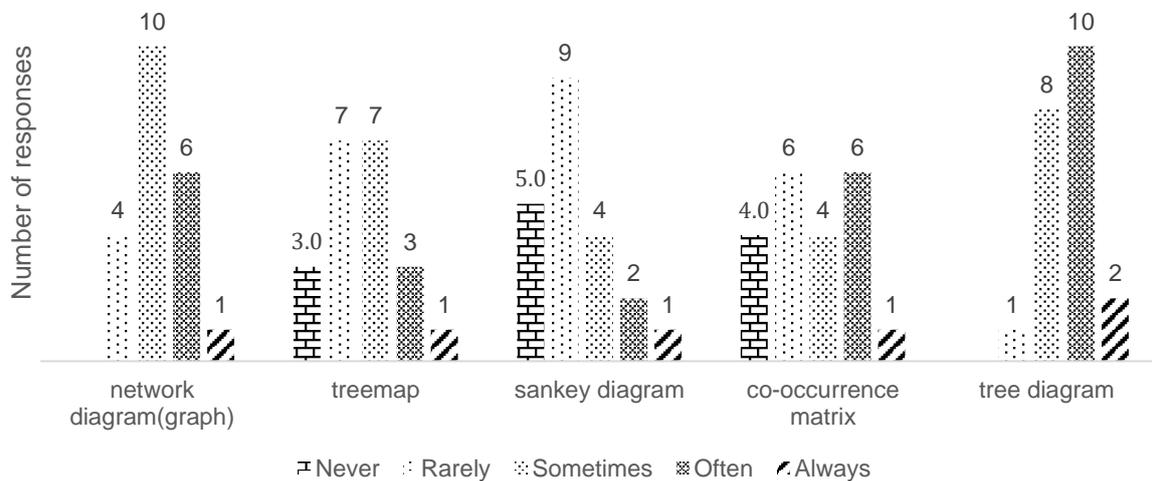
Figure 4. Familiarity with the five types of visualisations
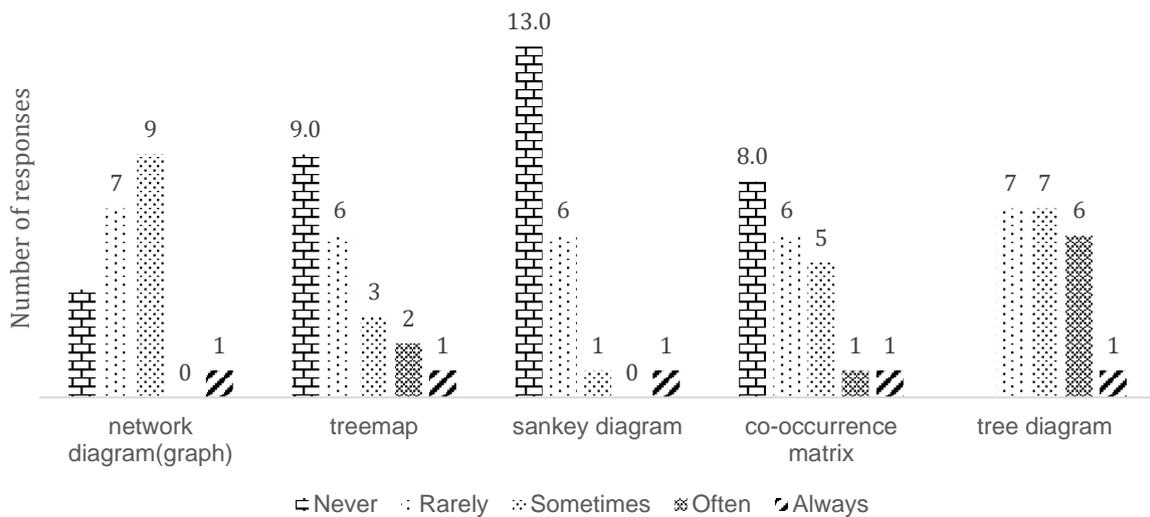
Figure 5. Work-related usage of visualisations

Moreover, as shown in Figure 6, calculated correlation coefficients between -0.3 and 0.14, suggest that, except for the co-occurrence matrix (correlation coefficient = 0.43, though at a statistically insignificant level with p-value = 0.08), there is no or weak correlation between prior familiarity with the provided visualisations and accuracy (i.e. participants' ability to determine the appropriate characteristic system' features of visualisations). One explanation for the inconclusive results on accuracy for the co-occurrence matrix is the effect of having administered the questionnaire on paper, with the matrix icon printed. Due to the low correlation coefficient, p-values are reported in the range between 0.07 and 0.96.

Likewise, correlation coefficients ranging from -0.09 to 0.26, suggest that accuracy of understanding the purpose of the visualisation and its work-related usage is not correlated. Again, similarly to the previous relationship between familiarity and accuracy, the co-occurrence matrix exhibits a moderate correlation (coefficient = 0.39), though, again, at a statistically insignificant level (p-value = 0.08).
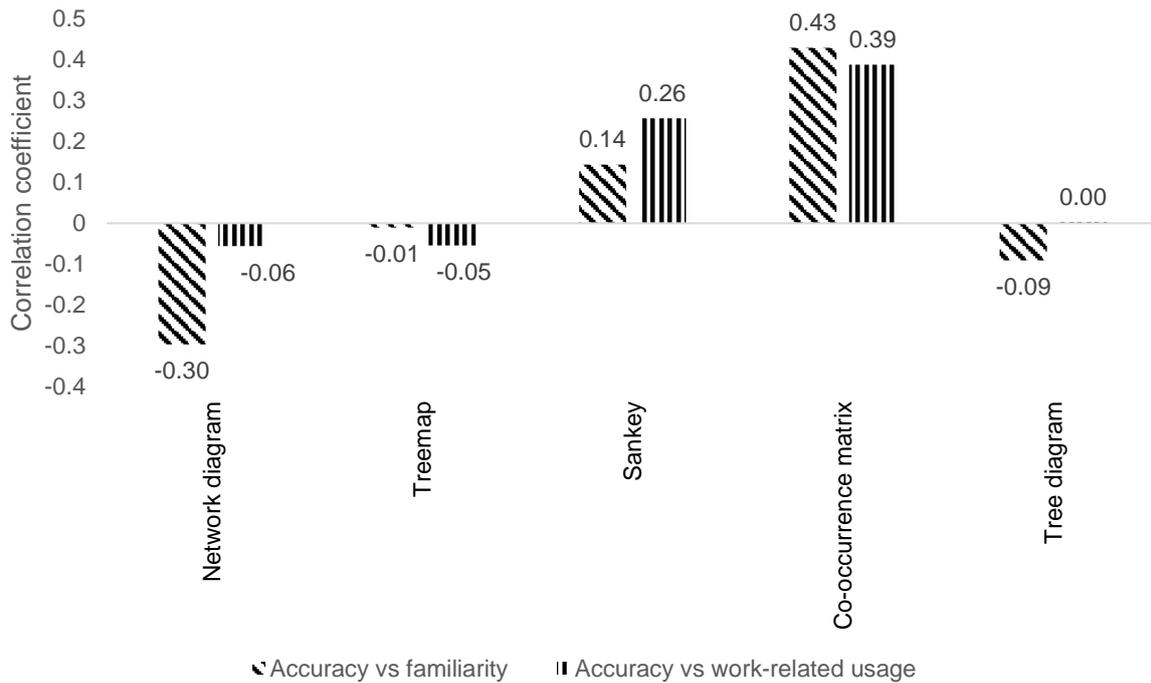
Figure 6. Correlation between familiarity, usage and accuracy

5 Discussion

On the whole, study participants were able to correctly identify characteristic system features represented by the selected information visualisations, which supports Q.1.1. When looking at the visualisations, the participants associated them with the characteristic system features that were previously defined as a ground truth. Moreover, by comparing the accuracy of detecting characteristic system features and previous familiarity with the visualisation, we found that the ability of the participants to correctly identify characteristic system features is not correlated with their previous familiarity with the visualisations, thereby rejecting Q.1.2.

While there are multiple factors that influence the sense-making process of unfamiliar visualisations (e.g. knowledge of content, data density, colour) (Lee et al., 2016), our results show that the provided information visualisations exhibit some inherent visual characteristics that help the participants understand characteristic features of systems underlying said visualisations: relationships between elements, hierarchies, patterns, processes. These findings are consistent with cognitive fit theory (Vessey, 1991) and the theory of affordances (Gibson, 1979). Firstly, through information visualisations, improved understanding of complex systems can be constructed among participants, and, secondly, the results indicate that information visualisations possess visual clues that allow users to understand their purpose.

An implication of the current study for product development, thus, is that training engineers to become more „data-savvy", including using and informing the design of information visualisations, can ultimately result in better products. Salustri et al. (2007) note that the choice of visualisation determines one's ability to recognise the characteristic features of systems from product information. Understanding these features during problem-solving in the early stages of design, for instance, then, leads to a higher probability of finding innovative solutions (Salustri et al., 2007). Therefore, knowing how to visualise characteristics systems' features, efficient visualisations that facilitate innovative thinking in product development can be prepared.

In addition, the visualisations in the study were presented on paper as opposed to being shown on a computer screen. We have observed that paper-based, digital visualisations printed on paper, is one of the common ways how these visualisations are used in practice. A systems engineer, for instance, would prepare a network diagram with the physical breakdown of a product, so that the team members are aligned on their tasks and know what exactly each part contains. Then, feedback from the team members can be quickly incorporated into the visualisation. Such an iterative process facilitates system overview for multiple design participants and elicits tailored discussions about system features.

6 Conclusions

Modern product development generates an abundance of data and information, which requires engineers to make sense of the data to make informed decisions. One way to represent the large amounts of diverse product information is through the use of information visualisations. Thus, this study contributes to the expanding research on the usage of novel visual techniques for improving understanding of complex product information.

In this paper, we have shown that data experts can accurately map system features to specific visualisations; in other words, recognise underlying features of complex systems – relationships, hierarchies, patterns and processes. The participants were able to identify characteristic systems' features represented by information visualisations with high accuracy, regardless of their previous familiarity with the visualisations. The results imply that training engineers to become more proficient in data and visual analysis could have benefits on understanding of product information, and, thus, product innovation.

Future research would include understanding the mapping between typical analytical tasks in product development, involved visual tasks and information visualisations that support these tasks. Then, through case-studies for different types of products (e.g. automotive and aerospace), such mapping could be evaluated. Finally, an empirical comparison with other visual representations used in product development is necessary.

Acknowledgements

References

Alfaris, A., Siddiqi, A., Rizk, C., De Weck, O. and Svetinovic, D. (2010), "Hierarchical Decomposition and Multidomain Formulation for the Design of Complex Sustainable Systems", *Journal of Mechanical Design*, American Society of Mechanical Engineers Digital Collection, Vol. 132 No. 9, pp. 0910031–09100313.

Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A. and Silva, C.T. (2011), "A User Study of Visualization Effectiveness Using EEG and Cognitive Load", *Computer Graphics Forum*, Blackwell Publishing Ltd, Vol. 30 No. 3, pp. 791–800.

Bracewell, R., Wallace, K., Moss, M. and Knott, D. (2009), "Capturing design rationale", *CAD Computer Aided Design*, Elsevier, Vol. 41 No. 3, pp. 173–186.

Card, S.K., Mackinlay, J. and Shneiderman, B. (1999), *Readings in Information Visualization: Using Vision to Think,* Morgan Kaufmann Publishers.

Clarkson, J., Simons, C. and Eckert, C. (2004), "Predicting change propagation in complex design", *Journal of Mechanical Design*, Vol. 126 No. 5, pp. 788–797.

Crawley, E., Cameron, B. and Selva, D. (2015), *System Architecture: Strategy and Product Development for Complex Systems*, Prentice Hall Press.

Danilovic, M. and Browning, T.R. (2007), "Managing complex product development projects with design structure matrices and domain mapping matrices", *International Journal of Project Management*, Vol. 25 No. 3, pp. 300–314.

Eppinger, S.D. and Browning, T.R. (2012), *Design Structure Matrix Methods and Applications*, *MIT Press Books*, Vol. 1.

Eppler, M.J. and Kernbach, S. (2016), "Dynagrams: Enhancing design thinking through dynamic diagrams", *Design Studies*, Elsevier, Vol. 47, pp. 91–117.

"EURITO: EU Relevant, Inclusive, Timely, Trusted, and Open Research Innovation Indicators". (2018), , available at: https://cordis.europa.eu/project/rcn/211945_en.html (accessed 11 July 2018).

Fawcett, T. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27 No. 8, pp. 861–874.

Heer, J., Bostock, M. and Ogievetsky, V. (2010), "A tour through the visualization zoo", *Communications of the ACM*, Vol. 53 No. 6, p. 59.

Hehenberger, P. (2014), "Perspectives on hierarchical modeling in mechatronic design", *Advanced Engineering Informatics*, Elsevier, Vol. 28 No. 3, pp. 188–197.

Idrissov, A., Skec, S. and Maier, A.M. (2020), "Visualising Systems: Mapping System Features and Interactive Information Visualisations in Design", *International Design Conference - DESIGN 2020*.

Inselberg, A. (1985), "The plane with parallel coordinates", *The Visual Computer*, Springer-Verlag, Vol. 1 No. 4, pp. 69–91.

Keller, R., Eckert, C.M. and Clarkson, P.J. (2006), "Matrices or node-link diagrams: Which visual representation is better for visualising connectivity models?", *Information Visualization*, Vol. 5 No. 1, pp. 62–76.

Lanzenberger, M., Miksch, S. and Pohl, M. (2005), "Exploring highly structured data a comparative study of stardinates and Parallel Coordinates", *Proceedings of the International Conference on Information Visualisation*, Vol. 2005, IEEE, pp. 312–320.

Larkin, J.H., Simon, H.A. and Simon, A. (1987), "Why a Diagramm is (Sometimes) Worth Ten Thousand Words", *Cognitive Science*, Vol. 11, pp. 65–99.

Lee, S., Kim, S.H., Hung, Y.H., Lam, H., Kang, Y.A. and Yi, J.S. (2016), "How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22 No. 1, pp. 499–508.

Norman, D. (2002), *The Design of Everyday Things*, MIT Press.

Novick, L.R. and Hurley, S.M. (2001), "To Matrix, Network, or Hierarchy: That Is the Question", *Cognitive Psychology*, available at:https://doi.org/10.1111/ajph.12083.

Pohl, M., Schmitt, M. and Diehl, S. (2009), "Comparing the Readability of Graph Layouts using Eyetracking and Task-oriented Analysis", *Computational Aesthetics in Graphics, Visualization, and Imaging*, pp. 49–56.

Riehmann, P., Hanfler, M. and Froehlich, B. (2005), "Interactive sankey diagrams", *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, pp. 233–240.

Shneiderman, B. (1992), "Tree visualization with tree-maps: 2-d space-filling approach", *ACM Transactions on Graphics*, Vol. 11 No. 1, pp. 92–99.

Sim, S.K. and Duffy, A.H.B.B. (2003), *Towards an Ontology of Generic Engineering Design Activities*, *Research in Engineering Design*, Vol. 14, available at:https://doi.org/10.1007/s00163-003-0037-1.

Spence, R. (2001), *Information Visualization*, New York: Addison-Wesley.

Star, S.L. and Griesemer, J.R. (1989), "Institutional Ecology , ' Translations ' and Boundary Objects : Amateurs and Professionals in Berkeley ' s Museum of Vertebrate Zoology", *Social Studies of Science*, Vol. 19 No. 3, pp. 387–420.

Stasko, J. and Zhang, E. (2000), "Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations", *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, Vol. 2000, pp. 57–65.

Steward, D. V. (1981), "Design Structure System: a Method for Managing the Design of Complex Systems.", *IEEE Transactions on Engineering Management*, Vol. EM-28 No. 3, pp. 71–74.

Summers, J.D. and Shah, J.J. (2010), "Mechanical engineering design complexity metrics: Size, coupling, and solvability", *Journal of Mechanical Design, Transactions of the ASME*, American Society of Mechanical Engineers Digital Collection, Vol. 132 No. 2, pp. 0210041–02100411.

Tufte, E.R. (2001), "The Visual Display of Quantitative Information, second edition.", *Technometrics*, Vol. 2nd, p. 197.

Vessey, I. (1991), "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature", *Decision Sciences*, Vol. 22 No. 2, pp. 219–240.

Yan, X., Qiao, M., Li, J., Simpson, T.W., Stump, G.M. and Zhang, X.L. (2012), "A Work-Centered Visual Analytics Model to Support Engineering Design with Interactive Visualization and Data-Mining", *2012 45th Hawaii International Conference on System Sciences*, IEEE, pp. 1845–1854.