



HAL
open science

Exploring the Potential of Virtual Reality for Model-Based Systems Architecting

Haobo Wang

► **To cite this version:**

Haobo Wang. Exploring the Potential of Virtual Reality for Model-Based Systems Architecting. Engineering Sciences [physics]. Université Grenoble Alpes [2020-..], 2024. English. NNT : 2024GRALI038 . tel-04718368

HAL Id: tel-04718368

<https://theses.hal.science/tel-04718368v1>

Submitted on 2 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : I-MEP2 - Ingénierie - Matériaux, Mécanique, Environnement, Energétique, Procédés, Production

Spécialité : GI - Génie Industriel : conception et production

Unité de recherche : Laboratoire des Sciences pour la Conception, l'Optimisation et la Production de Grenoble

Exploration du potentiel de la réalité virtuelle pour l'architecture système basée sur les modèles

Exploring the Potential of Virtual Reality for Model-Based Systems Architecting

Présentée par :

Haobo WANG

Direction de thèse :

Frédéric NOEL
PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes

Directeur de thèse

Romain PINQUIE
MAITRE DE CONFERENCES, Université Grenoble Alpes

Co-encadrant de thèse

Rapporteurs :

Ronan QUERREC
PROFESSEUR DES UNIVERSITES, École Nationale D'ingénieurs de Brest

Jean-Michel BRUEL
PROFESSEUR DES UNIVERSITES, Université de Toulouse

Thèse soutenue publiquement le **20 juin 2024**, devant le jury composé de :

Emilie LOUP-ESCANDE
PROFESSEURE DES UNIVERSITES, Université de Picardie Jules Verne

Présidente

Frédéric NOEL
PROFESSEUR DES UNIVERSITES, Centrale Supélec, LGI

Directeur de thèse

Ronan QUERREC
PROFESSEUR DES UNIVERSITES, Ecole Nationale d'ingénieurs de Brest

Rapporteur

Jean-Michel BRUEL
PROFESSEUR DES UNIVERSITES, Université de Toulouse

Rapporteur

Pierre DAVID
PROFESSEUR DES UNIVERSITES, Grenoble INP

Examineur

Guy-André BOY
PROFESSEUR DES UNIVERSITES, Centrale Supélec, LGI

Examineur

Invités :

Romain PINQUIE
MAITRE DE CONFERENCES, Université Grenoble Alpes



ABSTRACT

Model-based systems engineers and architects, particularly those moving from software to systems engineering, claim that SysML-like modelling notations, symbolic two-dimensional diagrams made of boxes and lines, are domain-independent and, thus, very convenient to support the cross-functional definition of a system architecture. However, the abstract diagramming syntax of Model-Based Systems Architecting (MBSA) notations makes their adoption difficult, especially by notational nonexperts, and using iconic graphics is one way of improvement. Few studies attempted to replace 2D diagrams with immersive 3D visuals without objective evidence. We assume it is due to limited development efforts and a need for more quality criteria for comparing 2D diagrams with 3D visuals. This thesis will argue that human-centric interactive 3D visuals should replace MBSA diagrams where appropriate to facilitate communication and participation in multidisciplinary co-design activities from mission to architecture definition. A combination of empirical validation methods shows that the proposed 3D immersive MBSA interface increases user satisfaction, provides better visual notations, and reduces cognitive load for single- and multi-user MBSA activities involving experts and non-experts. Such promising results of this exploratory research pave the way for more specific studies to cumulate scientific evidence.

Keywords: systems engineering, model-based systems engineering, systems architecting, model-based systems architecting, virtual reality, conceptual modelling language, human-computer interaction

ACKNOWLEDGEMENTS

The research presented in this thesis was carried out at the laboratory G-SCOP at INP Grenoble in France. It is part of the ArchiTOOL project funded by the French National Research Agency.

First, I want to thank my supervisor, Romain Pinquié, for his continuous support from the beginning till the end of this thesis. His frequent communication with him has greatly helped me advance, and he has led my way into the systems engineering world with his experienced opinions. Secondly, I want to thank my other supervisor, Frédéric Noël, a very experienced expert in virtual reality. Our exchange was less frequent, but his precise and refined advice was always a great help to my VR research. Both of my supervisors have been incredibly supportive and understanding, providing valuable research advice and offering emotional support during challenging times throughout my thesis. It was not easy at the beginning when there were still restrictions for the COVID-19 epidemic or when I injured my ankle and was stuck at home for several weeks. I am very lucky to have the support and guidance of both of my supervisors throughout my thesis journey. Without their assistance, I could never make it so far.

I want to thank my coworker, Sergio Camilo Medina Galvis, who helped me a lot with implementing the interface. I also want to thank my friends who supported me and shared their lives. The happy moments with them have been invaluable sources of strength to get me through this journey.

“The whole is greater than the sum of its parts.”

—Aristotle

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	ix
LIST OF TABLES	xv
LIST OF ABBREVIATIONS.....	xix
1 INTRODUCTION	21
1.1 Research Context	22
1.2 Research Problem	27
1.3 Research Questions and Hypotheses	30
1.4 Research Method.....	34
1.5 Statement of Contribution.....	35
1.6 Outline	38
2 LITERATURE REVIEW	39
2.1 Conceptual Modelling.....	39
2.1.1 Introduction to Conceptual Modelling.....	39
2.1.2 Graphical Conceptual Modelling Notations.....	42
2.1.3 Diagrammatic Reasoning	48
2.1.4 Quality of Graphical Conceptual Modelling Notations	57
2.1.5 Validation Method for Graphical Conceptual Modelling Notations.....	66
2.1.6 Discussion and Limits.....	71
2.2 Model-Based Systems Architecting	72
2.2.1 Systems Architecting.....	72
2.2.2 Model-Based Systems Architecting	75
2.2.3 Discussion and Limits.....	85
2.3 Immersive Human-Computer Interaction.....	85
2.3.1 Human-Computer Interaction	86
2.3.2 Immersive Human-Computer Interaction	87
2.3.3 Immersive Human-Computer Interaction for Software Architecting.....	90
2.3.4 Immersive Human-Computer Interaction for MBSA	93
2.3.5 Discussion and Limits.....	96
2.4 Validation of Immersive Human-Computer Interaction.....	97
2.4.1 Validation Method of Human-Computer Interaction	97

2.4.2	Quality Criteria of Human-Computer Interaction	100
2.4.3	Discussion and Limits.....	104
2.5	Overall Discussion and Limits	104
3	IMMERSIVE MODEL-BASED SYSTEMS ARCHITECTING INTERFACE.....	105
3.1	MBSA Engineer Tasks Analysis	105
3.1.1	Tasks from ISO 15288	105
3.1.2	Tasks from ISO 24641	106
3.1.3	Synthesis of Systems Architect Tasks from ISO 15288 and 24641	108
3.1.4	Tasks from Existing MBSA Methods.....	109
3.1.5	MBSA Tasks Selection.....	110
3.1.6	MBSA Data Model.....	111
3.1.7	MBSA Modelling Method.....	113
3.2	Architecture.....	114
3.3	Visual Metaphors and Interactions	114
4	EMPIRICAL VALIDATION	127
4.1	Validation Method	127
4.2	Selection of Quality Criteria for Evaluation of Immersive MBSA Interface... ..	128
4.2.1	Measure of User Satisfaction	130
4.2.2	Measure of Visual Notation	131
4.2.3	Measure of Presence	131
4.2.4	Measure of Cognitive Load.....	132
4.2.5	Measure of Group Dynamics.....	132
4.3	Online Survey	132
4.3.1	Research Questions.....	132
4.3.2	Online Survey Design.....	134
4.3.3	Online Survey Validation	136
4.3.4	Online Survey Diffusion	139
4.3.5	Online Survey Analysis	140
4.3.6	Online Survey Results.....	142
4.3.7	Online Survey Discussion.....	152
4.3.8	Threats to Validity.....	154
4.4	Single-Person Controlled Experiment.....	154
4.4.1	Research Questions.....	154
4.4.2	Single-Person Controlled Experiment Participants	155

4.4.3 Single-Person Controlled Experiment Design.....	155
4.4.4 Single-Person Controlled Experiment Validation	157
4.4.5 Single-Person Controlled Experiment Analysis.....	158
4.4.6 Single-Person Controlled Experiment Results	159
4.4.7 Single-Person Controlled Experiment Discussion.....	164
4.5 Multi-Person Controlled Experiment.....	165
4.5.1 Research Questions.....	165
4.5.2 Multi-Person Controlled Experiment Participants.....	166
4.5.3 Multi-Person Controlled Experiment Design	166
4.5.4 Multi-Person Controlled Experiment Validation.....	168
4.5.5 Multi-Person Controlled Experiment Analysis	169
4.5.6 Multi-Person Controlled Experiment Results	169
4.5.7 Multi-Person Controlled Experiment Discussion	174
4.6 Formative Usability Tests with MBSA Experts	174
4.6.1 Research Questions.....	174
4.6.2 Formative Usability Testing Participants.....	175
4.6.3 Formative Usability Testing Design	176
4.6.4 Formative Usability Testing Validation.....	178
4.6.5 Formative Usability Testing Analysis	178
4.6.6 Formative Usability Testing Results	178
4.6.7 Formative Usability Testing Discussion	184
4.7 Meta-Analysis	185
4.7.1 User Satisfaction	185
4.7.2 Visual Notation	190
4.7.3 Presence.....	195
4.7.4 Cognitive Load	198
4.7.5 Summary.....	198
5 DISCUSSION	201
5.1 Answers to Research Questions	201
5.2 Threats to Validity	202
5.3 Reflections on More Generic Questions	203
5.3.1 How do 3D representations differ from diagrammatic ones concerning systems architecting tasks?	203
5.3.2 How did the immersive MBSA interface help architects complete their tasks?	204

5.3.3 Is there a difference between notational experts and novices?	206
5.3.4 What barriers did users face when using the immersive MBSA interface?	206
5.3.5 Did the immersive MBSA interface facilitate co-design activities?	207
5.3.6 Could people imagine using a similar immersive MBSA interface?	207
5.3.7 How would user improve the MBSA interface to integrate it into their work? .	208
6 CONCLUSION	211
6.1 Contribution and Claim	211
6.2 Perspectives: A Vision of VR for MBSA	212
REFERENCES	215
APPENDICES	227

LIST OF FIGURES

Figure 1 The diversity of visual external representations to communicate a system architecture (Formentini et al. 2022; Grundel and Abulawi 2016; Hause 2011; Mhenni et al. 2014; Weilkiens 2020)	21
Figure 2 Illustrations of SysML diagrams (Hernandez and Fernandez-Sanchez 2017; Hoffmann 2014; Mhenni et al. 2014).....	23
Figure 3 Interrelationships between processes (ISO/IEC/IEEE 15288 2023).....	24
Figure 4 Model-Based Systems and Software Engineering (MBSSE) Reference model (ISO/IEC/IEEE 24641 2023)	25
Figure 5 Overview of visual representations, cognition process, and activity performance ...	27
Figure 6 Conventional modelling scenario and VR technology-based modelling scenario	30
Figure 7 Overview of the research method	35
Figure 8 MBSE in a slide (Holt 2021).....	36
Figure 9 Expanding "MBSE in a slide" to include implementation (Holt 2021)	37
Figure 10 The FBS framework after (Gero and Kannengiesser 2004). Processes: (1) formulation, (2) synthesis, (3) analysis, (4) evaluation, (5) documentation. Reformulation processes, which are included in the original version of this figure, have been omitted for clarity (Müller 2020)	40
Figure 11 Design process with embodiment design (Langeveld 2011)	40
Figure 12 Product functional decomposition allocated to logical subsystems (left) and architecture geometrical definition (right)	41
Figure 13 Illustration of Visual notation.....	43
Figure 14 Representation of a tree at various levels of abstraction (Grundel and Abulawi 2016)	44
Figure 15 Saussure's model of the sign (Chandler 2007)	44
Figure 16 Peirce's semiotic triangle (Chandler 2007) on the left and sign elements and modes in semiotics (Kuhar and Polančič 2021) on the right	45
Figure 17 Iconic ERD (left) and Standard ERD (right) (Masri, Parker, and Gemino 2008)	46
Figure 18 Example of SkiPo model and elaboration of design details (Grundel and Abulawi 2016).....	47
Figure 19 An example of a geon diagram and its UML equivalent (Irani and Ware 2003).....	47
Figure 20 Illustration of stereoscopic display (Wai-keung Fung et al. 2005).....	48
Figure 21 Modification for Visual Cues with Class Diagram (Top), Event Trace Diagram (Left), and Object Message Diagram (Right) (Kim, Hahn, and Hahn 2000)	51
Figure 22 Diagram Transition Graph of Control Group (Left) and Experimental Group (Right) (Kim, Hahn, and Hahn 2000)	52
Figure 23 Modification for Contextual Information with Use Case Diagram (Left) and Context Diagram (Right) (Kim, Hahn, and Hahn 2000)	52
Figure 24 Hypothesis Behavior Graph of Control Group (Left) and Experimental Group (Right) (Kim, Hahn, and Hahn 2000)	53

Figure 25 The Cognitive Theory of Multimedia Learning (Mayer 2001).....	54
Figure 26 Illustration of the diagrammatic reasoning	56
Figure 27 Illustration with the visual system.....	56
Figure 28 Scope of PoN on the top left-hand quadrant of the diagram (Moody 2009).....	60
Figure 29 Anomalies of Semiotic Clarity (Moody 2009)	60
Figure 30 Bertin's eight visual variables (Moody 2009).....	61
Figure 31 Semantically transparent relationships (Moody 2009).....	62
Figure 32 Hierarchical organisation (Moody 2009)	62
Figure 33 Levels of visual expressiveness: (a) UML=0, (b) IE=1 (shape), and (c) Oracle=2 (shape, brightness) (Moody 2009)	63
Figure 34 Illustration of dual coding (Moody 2009)	63
Figure 35 Interactions between principles (Moody 2009).....	64
Figure 36 Symbol redundancies (left) and Symbol overload (right) (Anwer and El-Attar 2014)	68
Figure 37 Comparison of a circle shape and Rectangle shape (Anwer and El-Attar 2014) ...	69
Figure 38 Equations to calculate the metrics of Semiotic Clarity.....	70
Figure 39 Diagrams used in the user evaluation (Figl and Derntl 2011).....	71
Figure 40 System lifecycle processes (left) and the scope of this thesis (right) (ISO/IEC/IEEE 15288 2023)	74
Figure 41 Conceptual model of an architecture description framework (ISO/IEC/IEEE 42010 2022)	75
Figure 42 Conceptual model of an architecture description language (ISO/IEC/IEEE 42010 2022)	75
Figure 43 ISE Subprocess (Fernandez and Hernandez 2019).....	77
Figure 44 PPOOA Subprocess (Fernandez and Hernandez 2019).....	77
Figure 45 Analyzing requirements (Weilkiens 2020).....	78
Figure 46 Architecture Design	78
Figure 47 Overview of Harmony-SE (Forsberg and Mooz 1991)	79
Figure 48 Overview of SE-READ (Mhenni et al. 2014)	80
Figure 49 Overview of ASAP (Guillermo and Góngora 2014)	81
Figure 50 Overview of SysCARS (J. Piques 2014)	81
Figure 51 Overview of ARCADIA (Voirin 2018).....	82
Figure 52 Overview of SysML/UML interrelationship (OMG 2019).....	83
Figure 53 SysML diagram taxonomy (OMG 2019).....	83
Figure 54 GUI of OPCAT shows the baking system's (SD, top-level) system diagram. Top: object–process diagram (OPD). Bottom: The corresponding, automatically generated Object–Process Language (OPL) paragraph (Dori 2011).....	84
Figure 55 A top-down view of the semantic network of the human as an information processor (Dix 2003).....	86
Figure 56 A framework for human-computer interaction (Dix 2003).....	87

Figure 57 Simplified representation of a "virtuality continuum" (Milgram and Kishino 1994)..	88
Figure 58 Classification of VR based on the type of technology used to achieve virtual worlds and interact with them (Kommetter and Ebner 2019)	88
Figure 59 Navigation techniques (Bowman, Koller, and Hodges 1997) (left), manipulation and selection techniques (Bowman, Johnson, and Hodges 2001) (right)	89
Figure 60 Classification of VE manipulation techniques depending on their underlying metaphors (Poupyrev 1998)	89
Figure 61 Taxonomy of characteristics for navigation, selection and manipulation (Raimbaud 2020)	90
Figure 62 Island metaphor (Schreiber et al. 2019)	90
Figure 63 Example of visualisation (Maletic et al. 2001)	91
Figure 64 An UML Class Diagram example with custom 3D representation (Yigitbas et al. 2022)	91
Figure 65 An example of VR-UML (Oberhauser 2021)	92
Figure 66 Overview of VR-MF (Oberhauser 2021)	92
Figure 67 Overview of the complete process for VR scene configuration (Mahboob et al. 2017)	93
Figure 68 Examples of the visualisation method (Lutfi and Valerdi, 2023)	94
Figure 69 3D visualisation of a filling station with a 3D State Machine Diagram (Barosan and Van Der Heijden 2022)	94
Figure 70 Overview of CONOPS: Screenshot of CES user interface (top) and two CES scenarios (bottom) (Korfiatis, Cloutier, and Zigh 2012)	95
Figure 71 Conceptual map of various VR solution concepts in VR-MF (Oberhauser 2022) ..	96
Figure 72 Multiple and heterogeneous side-by-side models in VR (Oberhauser 2022).....	96
Figure 73 Overview of analytic evaluation techniques (Dix 2003)	97
Figure 74 Overview of experimental and query evaluation techniques (Dix 2003)	98
Figure 75 Overview of observational evaluation techniques (Dix 2003)	98
Figure 76 Overview of monitoring evaluation techniques (Dix 2003).....	99
Figure 77 Methodology for the user-centred design and evaluation of VE user interaction (Gabbard, Hix, and Swan 1999).....	100
Figure 78 Relationship between human factors, shown at the centre, and other related disciplines of study (Wickens et al. 2004)	101
Figure 79 List of Ergonomic Criteria for HVEIs (Bach and Scapin 2003).....	102
Figure 80 Overview of usability dimensions (Di Gironimo et al. 2013)	103
Figure 81 SUS score scale (Bangor et al. 2009).....	103
Figure 82 Illustrations of the 2D diagrammatic MBSA interface (Mhenni et al. 2014) and 3D immersive MBSA interface.....	105
Figure 83 The relationships among MBSSE process groups and build models processes (ISO/IEC/IEEE 24641 2023)	107
Figure 84 MBSSE reference model (ISO/IEC/IEEE 24641 2023)	107
Figure 85 Data model of the 3D immersive MBSA interface.	112

Figure 86 Overview of the method.....	113
Figure 87 Overview of the architecture	114
Figure 88 Illustration of buttons on the controller	115
Figure 89 Illustration of ray-casting.....	115
Figure 90 Main menu content at lifecycle definition tasks (left) and context definition task (right)	116
Figure 91 Contextual menu for a stakeholder (left) and the system of interest seen as a black box (right)	116
Figure 92 Navigation menu	117
Figure 93 System lifecycle.....	117
Figure 94 Example of creating a system context	118
Figure 95 System context.....	118
Figure 96 System with main menu (left) and contextual menu (right).....	119
Figure 97 External entities library from the main menu	119
Figure 98 Requirement editing view on a 2D screen with the mouse and keyboard devices (top) and the requirement template (bottom).....	120
Figure 99 Interface with three flows (top) and a flow exchanged between an output red port and an input blue port (bottom). Textual statements attached to the flows correspond to the definition of functions.	121
Figure 100 Interfaces (Formentini et al. 2022)	122
Figure 101 Function (play voice information button)	122
Figure 102 Example of operational scenario diagram for an electronic toothbrush (Krob 2022)	123
Figure 103 Scenario	123
Figure 104 Illustration from inside of black box.....	124
Figure 105 CSG Tree with the three Boolean operations: union \cup , intersection \cap , difference \setminus on primitives (Wassermann et al. 2016).....	124
Figure 106 Example of subtract.....	124
Figure 107 Illustration of facet selection (left) and polygon (right)	125
Figure 108 Subsystem.....	125
Figure 109 Photos of experimentation	126
Figure 110 Illustration of spatial relation	133
Figure 111 Capture of the online survey demographic question	134
Figure 112 Representation 1	135
Figure 113 Representation 2	135
Figure 114 Representation 3	135
Figure 115 Representation 4	135
Figure 116 Representation 5	135
Figure 117 Representation 6	135
Figure 118 Representation 7	135

Figure 119 The relationship between reliability and validity is illustrated as “target diagrams” (Lundberg 2006)	137
Figure 120 Subtypes of various forms of validity tests (Taherdoost 2016)	137
Figure 121 Boxplot of the Perceptual Discriminability result.....	143
Figure 122 Boxplot of the Semantic Transparency result.....	145
Figure 123 Boxplot of the Complexity Management result.....	147
Figure 124 Boxplot of the Visual Expressiveness result.....	148
Figure 125 Figure 15 Boxplot of the Graphic Economy result	150
Figure 126 Pie chart on the percentage of the answers (2D left, 3D right)	152
Figure 127 Set up of a single-person controlled experiment	156
Figure 128 Example of "Presence profile" from study PQII	158
Figure 129 IPQ result	163
Figure 130 IPQ scores (3D immersive MBSA interface scores and mean scores)	164
Figure 131 Set up of a multi-person controlled experiment.....	167
Figure 132 Photo of formative usability testing	177
Figure 133 SUS scores	179
Figure 134 Perceived efficiency.....	179
Figure 135 Perceived effectiveness.....	179
Figure 136 Perceived retention rate.....	179
Figure 137 Intention to use in the future	179
Figure 138 Usage fun	179
Figure 139 User motivation	180
Figure 140 Perceptual discriminability	180
Figure 141 Semantic transparency	180
Figure 142 Complexity management.....	180
Figure 143 Visual expressiveness	180
Figure 144 Graphic economy	181
Figure 145 Presence	181
Figure 146 Spatial presence.....	181
Figure 147 Involvement.....	181
Figure 148 Experienced realism	181
Figure 149 NASA-TLX raw score	182
Figure 150 Overview of SUS score.....	186
Figure 151 Overview of Perceived Efficiency.....	186
Figure 152 Overview of Perceived Effectiveness.....	187
Figure 153 Overview of Perceived Retention Rate	187
Figure 154 Overview of Intention to use in the future.....	188
Figure 155 Overview of Usage Fun	188

Figure 156 Overview of User Motivation	189
Figure 157 Survey result of Perceptual Discriminability	190
Figure 158 Summary of Perceptual Discriminability	191
Figure 159 Survey result of Semantic Transparency	191
Figure 160 Summary of Semantic Transparency	192
Figure 161 Survey result of Complexity Management	192
Figure 162 Summary of Complexity Management	193
Figure 163 Survey result of Visual Expressiveness	193
Figure 164 Summary of Visual Expressiveness	194
Figure 165 Survey result of Graphic Economy	194
Figure 166 Summary of Graphic Economy	195
Figure 167 Summary of Presence	196
Figure 168 Summary of Spatial Presence	196
Figure 169 Summary of Involvement	197
Figure 170 Summary of Experienced Realism	197
Figure 171 Summary of cognitive load	198
Figure 172 Bertin's eight visual variables (D. Moody 2009)	204
Figure 173 Illustration of diagrammatic reasoning with visual system	205
Figure 174 Overview of Perceived Retention Rate	206

LIST OF TABLES

Table 1 Gist of the thesis.....	26
Table 2 Seven principles of CTML (Masri, Parker, and Gemino 2008)	54
Table 3 Properties of visual variables (Moody 2009)	63
Table 4 Comparison between Principles for modelling language design and PoN.....	65
Table 5 Comparisons between CDs and PoN	65
Table 6 Review of the main SysML methods to identify the concepts and representations of an operational view.....	110
Table 7 Concepts in 3D immersive MBSA interface	111
Table 8 Validation methods and research questions	127
Table 9 Overview table of criteria	129
Table 10 Validation methods and research questions (Survey)	133
Table 11 Representation and PoN Criteria Mapping.....	136
Table 12 Results of Cronbach's Alpha for each PoN criteria.....	139
Table 13 Representation and PoN Criteria Mapping.....	141
Table 14 Within-Subjects Factors of ANOVA test for Perceptual Discriminability.....	142
Table 15 Descriptive Statics of ANOVA test for Perceptual Discriminability	142
Table 16 Multivariate Tests of ANOVA test for Perceptual Discriminability	143
Table 17 Pairwise Comparisons of ANOVA test for Perceptual Discriminability.....	143
Table 18 Within-Subjects Factors of ANOVA test for Semantic Transparency.....	144
Table 19 Descriptive Statistics of ANOVA test for Semantic Transparency	144
Table 20 Multivariate Tests of ANOVA test for Semantic Transparency	144
Table 21 Pairwise Comparisons of ANOVA test for Semantic Transparency.....	145
Table 22 Within-Subjects Factors of ANOVA test for Complexity Management.....	145
Table 23 Descriptive Statistics of ANOVA test for Complexity Management.....	146
Table 24 Multivariate Tests of ANOVA test for Complexity Management.....	146
Table 25 Pairwise Comparisons of ANOVA test for Complexity Management	146
Table 26 Within-Subjects Factors of ANOVA test for Visual Expressiveness.....	147
Table 27 Descriptive Statistics of ANOVA test for Visual Expressiveness	147
Table 28 Multivariate Tests ^a of ANOVA test for Visual Expressiveness	147
Table 29 Pairwise Comparisons of ANOVA test for Visual Expressiveness.....	148
Table 30 Within-Subjects Factors of ANOVA test for Graphic Economy.....	148
Table 31 Descriptive Statistics of ANOVA test for Graphic Economy.....	149
Table 32 Multivariate Tests ^a of ANOVA test for Graphic Economy	149
Table 33 Pairwise Comparisons of ANOVA test for Graphic Economy	149
Table 34 Results of independent sample t-test.....	150

Table 35 Paired Samples Statistics of paired sample t-test for spatial relation	151
Table 36 Paired Samples Test of paired sample t-test for spatial relation.....	151
Table 37 Validation methods and research questions (Expe_S).....	155
Table 38 Factors of independent sample t-test.....	159
Table 39 Group statistics of independent sample t-test for SUS	159
Table 40 Group statistics of independent sample t-test for perceived efficiency	160
Table 41 Group statistics of independent sample t-test for perceived effectiveness	160
Table 42 Group statistics of independent sample t-test for perceived retention rate	160
Table 43 Group statistics of independent sample t-test for intention to use in the future.....	160
Table 44 Group statistics of independent sample t-test for usage fun.....	161
Table 45 Group statistics of independent sample t-test for user motivation	161
Table 46 Group statistics of independent sample t-test for perceptual discriminability	161
Table 47 Group statistics of independent sample t-test for semantic transparency.....	162
Table 48 Group statistics of independent sample t-test for complexity management.....	162
Table 49 Group statistics of independent sample t-test for visual expressiveness	162
Table 50 Group statistics of independent sample t-test for graphic economy	162
Table 51 Descriptive statistics for four factors of IPQ.....	163
Table 52 Group statistics of independent sample t-test for NASA-TLX raw score.....	164
Table 53 Validation methods and research questions (Expe_M)	166
Table 54 Paired samples statistics of paired sample t-test for SUS	169
Table 55 Paired samples statistics of paired sample t-test for perceived efficiency	169
Table 56 Paired samples statistics of paired sample t-test for perceived effectiveness.....	170
Table 57 Paired samples statistics of paired sample t-test for perceived retention rate	170
Table 58 Paired samples statistics of paired sample t-test for intention to use in the future	170
Table 59 Paired samples statistics of paired sample t-test for usage fun.....	170
Table 60 Paired samples statistics of paired sample t-test for user motivation	171
Table 61 Paired samples statistics of paired sample t-test for perceptual discriminability ...	171
Table 62 Paired samples statistics of paired sample t-test for semantic transparency.....	171
Table 63 Paired samples statistics of paired sample t-test for complexity management	172
Table 64 Paired samples statistics of paired sample t-test for visual expressiveness	172
Table 65 Paired samples statistics of paired sample t-test for graphic economy.....	172
Table 66 Paired samples statistics of paired sample t-test for NASA-TLX raw score.....	172
Table 67 Paired samples statistics of paired sample t-test for perceived efficiency of communication	173
Table 68 Paired samples statistics of paired sample t-test for perceived effectiveness of communication	173
Table 69 Paired samples statistics of paired sample t-test for member activeness.....	173
Table 70 Validation methods and research questions (Formative).....	175

Table 71 Group with expertise level.....	176
Table 72 Validation methods and research questions (RQ1)	185
Table 73 Validation methods and research questions (RQ2)	190
Table 74 Validation methods and research questions (RQ3)	195
Table 75 Validation methods and research questions (RQ4)	198
Table 76 Systems Principles (Rick et al. 2023).....	227
Table 77 Activities and tasks of business or mission analysis process (ISO/IEC/IEEE 15288 2023)	229
Table 78 Activities and tasks of stakeholder needs and requirements definition process (ISO/IEC/IEEE 15288 2023)	229
Table 79 Activities and tasks of system requirements definition process (ISO/IEC/IEEE 15288 2023)	230
Table 80 Activities and tasks of system architecture definition process (ISO/IEC/IEEE 15288 2023)	231
Table 81 Activities and tasks of produce system models process (ISO/IEC/IEEE 24641 2023)	233
Table 82 Activities and tasks of perform business and mission analysis process (ISO/IEC/IEEE 24641 2023)	233
Table 83 Activities and tasks of perform operational analysis process (ISO/IEC/IEEE 24641 2023)	234
Table 84 Activities and tasks of perform functional analysis process (ISO/IEC/IEEE 24641 2023)	235
Table 85 Activities and tasks of perform system structure design process (ISO/IEC/IEEE 24641 2023)	235

LIST OF ABBREVIATIONS

SA	Systems Architecting
SE	Systems Engineering
MBSA	Model-Based Systems Architecting
MBSE	Model-Based Systems Engineering
MBSSE	Model-Based Systems and Software Engineering
INCOSE	International Council on Systems Engineering
ISO	International Organization for Standardization
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
VR	Virtual Reality
VE	Virtual Environment
HMD	Head-Mounted Display
SysML	System Modelling Language
UML	Unified Modelling Language
OPM	Object-Process Methodology
HCI	Human-Computer Interaction
HVEI	Human-Virtual environment interactions
ERD	Entity-Relationship Diagrams
BPMN	Business Process Model and Notation
CLT	Cognitive Load Theory
CTML	Cognitive Theory of Multimedia Learning
PDK	Previous Domain Knowledge
ESL	English as a Second Language
CDs	Cognitive Dimensions of Notations
PoN	Physics of Notations
NASA	National Aeronautics and Space Administration
CONOPS	Concept of Operations
SUS	System Usability Scale
CSG	Constructive Solid Geometry

1 INTRODUCTION

In the design world, representations are essential in delivering ideas and facilitating communication between stakeholders involved in designing complex engineered systems. The transdisciplinary and integrative dimensions of systems engineering require stakeholders to share and understand each other viewpoints as they have different concerns. Representations serve to connect abstract concepts with tangible instances. According to (Visser 2010), at a cognitive level, the design consists of constructing (generating, transforming, and evaluating) partial representations of a chosen part of an artefact product until they are so precise, concrete, and detailed that the resulting representations specify explicitly and completely the implementation of the artefact product. Linguistic, graphical, gestural, [...], and postural representations mainly serve to communicate design information among members of an interdisciplinary design team and with external stakeholders throughout the design process. Representations can be “mental cognitive artefacts” (e.g., ideas) or superficial (e.g., verbal descriptions, sketches, 3D models, etc.). In this research, we concentrate on visual external representations created and utilised by systems architects to communicate and validate the architecture of engineered systems.

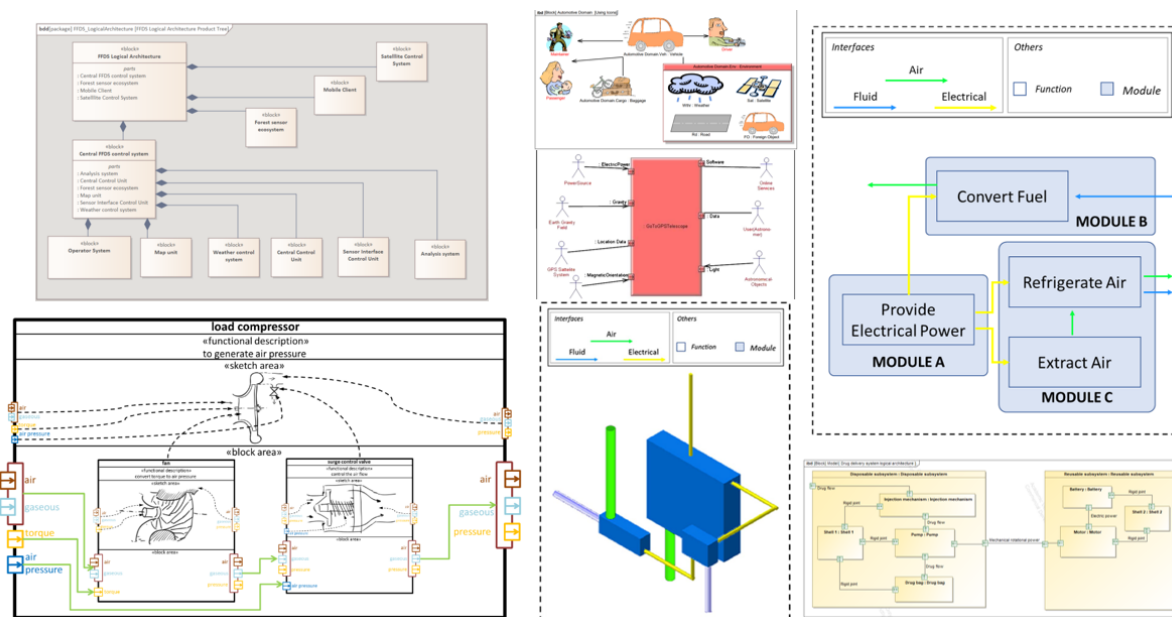


Figure 1 The diversity of visual external representations to communicate a system architecture (Formentini et al. 2022; Grundel and Abulawi 2016; Hause 2011; Mhenni et al. 2014; Weilkiens 2020)

At the core of the ArchiTOOL project funded by the French National Research Agency (ANR), this thesis aims to invent, prototype, and evaluate new immersive and interactive modelling interfaces for architecting engineered systems to advance human-centred

computing in systems engineering. As exploratory research, we investigate the potential of using virtual reality to co-design the architecture of engineered systems. Starting from such a general idea, this pioneer investigation intends to understand better the unclear research hypothesis on the use of virtual reality for systems architecting and identify more specific issues that can be the focus of future research rather than offering final and conclusive solutions to existing well-defined issues. Despite this exploratory dimension, we will adopt the principles of empirical research by comparing our immersive modelling interface with a conventional diagrammatic model-based systems engineering interface.

1.1 Research Context

This thesis will concentrate on the systems engineering approach among the various logical ways to design a system (e.g., design thinking, design for Six Sigma, etc.). The first time the term systems engineering was brought up can be traced back to Bell Telephone Laboratories in the 1940s for the need to identify and manipulate the properties of a system as a whole, which implies that we may encounter situations where the properties of a whole system can be significantly different from the properties of the sum of the parts. Those needs motivated industries, especially the US military in the late 1940s with the initial development of missiles and missile-defence systems, to apply the systems engineering approach.

Systems engineering (SE) mainly focuses on architecture definition. System architecture definition is the process of generating system architecture alternatives, selecting one or more alternative(s) that address stakeholder concerns and system requirements, and expressing this in consistent views (e.g., business, operational, functional, behavioural, logical, structural) and models (ISO/IEC/IEEE 15288 2023). A paradigm shift from document-centric to model-based has occurred during the evolution of systems engineering, bringing up a new concept supporting the architecture definition process: Model-Based Systems Engineering (MBSE) approach. MBSE is the systematic and formalised application of modelling to support SE (Engineering 2015).

MBSE has been applied to numerous industrial projects in various domains, from defence and environment to transportation, automotive, and aerospace. For instance, the National Aeronautics and Space Administration (NASA) has employed MBSE in numerous space missions, including the Mars Rover missions and the development of the Space Launch System (SLS) (Holladay et al. 2019). In the automotive industry, Ford has embraced MBSE to streamline the design and development of vehicles (Biggs et al., 2018). Among these projects, the domain-specified methodologies are constructed with practical experiments for better adoption since the requirements and specifications differ across disciplines.

With or without some domain-specific tailored modification, the System Modelling Language (SysML) is often deployed as an MBSE language. In the 2D SysML diagrams, the models of a real object vary from very abstract symbolic models to very concrete iconic models (Grundel and Abulawi 2016), which strive to facilitate communication among all stakeholders involved in the system architecture definition process.

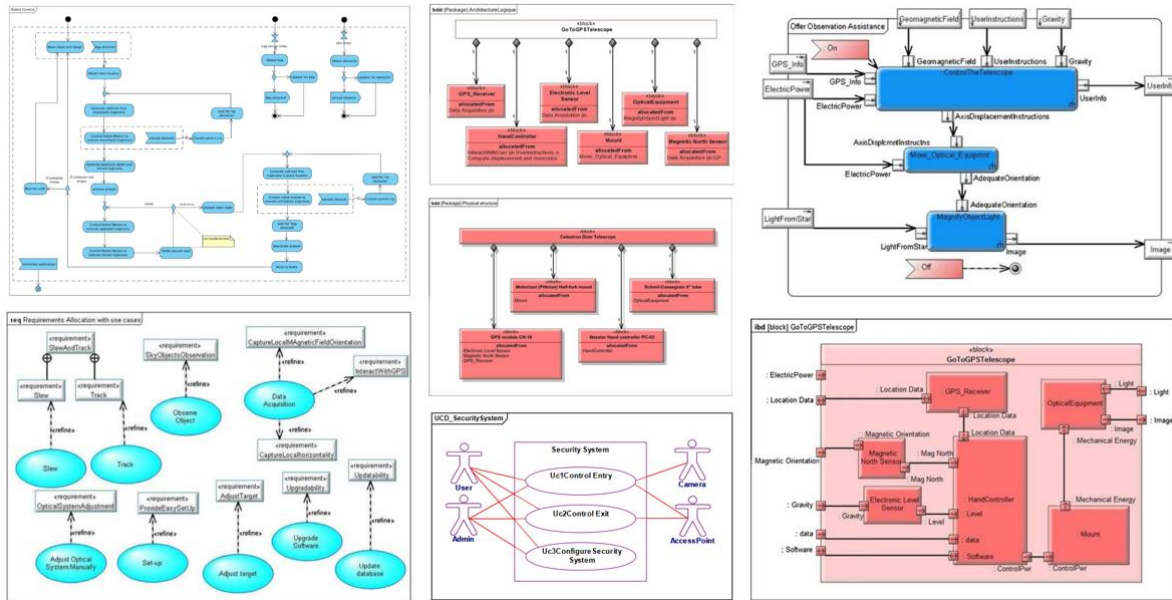


Figure 2 Illustrations of SysML diagrams (Hernandez and Fernandez-Sanchez 2017; Hoffmann 2014; Mhenni et al. 2014)

Apart from the 2D diagrams, researchers also try to build immersive environments to improve the performance of the MBSE approach. For instance, there exist attempts to develop graphical concepts of operation leveraging gaming technology, which can provide system developers with direct access to the needs of stakeholders (Korfiatis, Cloutier, and Zigh 2012) or to visualise a software architecture in virtual reality using an island metaphor (Schreiber and Misiak 2018). Using concrete graphical 3D representation rather than abstract symbolic 2D diagrams intends to increase the figurative tangibility of MBSE concepts. As presented in (Boy 2017) physical tangibility is “*the property of an object that is physically graspable (i.e., you can touch it, hold it, sense it and so on)*”, and figurative tangibility is “*the property of a concept that is cognitively graspable (i.e., you can understand it, appropriate it, feel it and so on)*”.

In the meantime, it is crucial to find an efficient way of assessing and evaluating all these nascent systems' architecting modelling capabilities, admitting that there are existing approaches to evaluate the usability and interactivity of an immersive environment or the construction of visual notations. Still, no single method can systematically cover the full range of criteria for comparing the diversity of external visual representations (2D/3D,

iconic/symbolic, monoscopic/stereoscopic, etc.) to communicate systems architecting information among members of an interdisciplinary new product development team.

Last but not least, we want to clarify the scientific position of this thesis in the related domain. System Cycle engineering practitioners may have terminology that sometimes differs from others depending on their own experiences and the communities they belong to (e.g., engineering design or systems engineering). As said (Holt 2021), one aspect of all engineering is that there is seldom a single, definitive definition for any term.

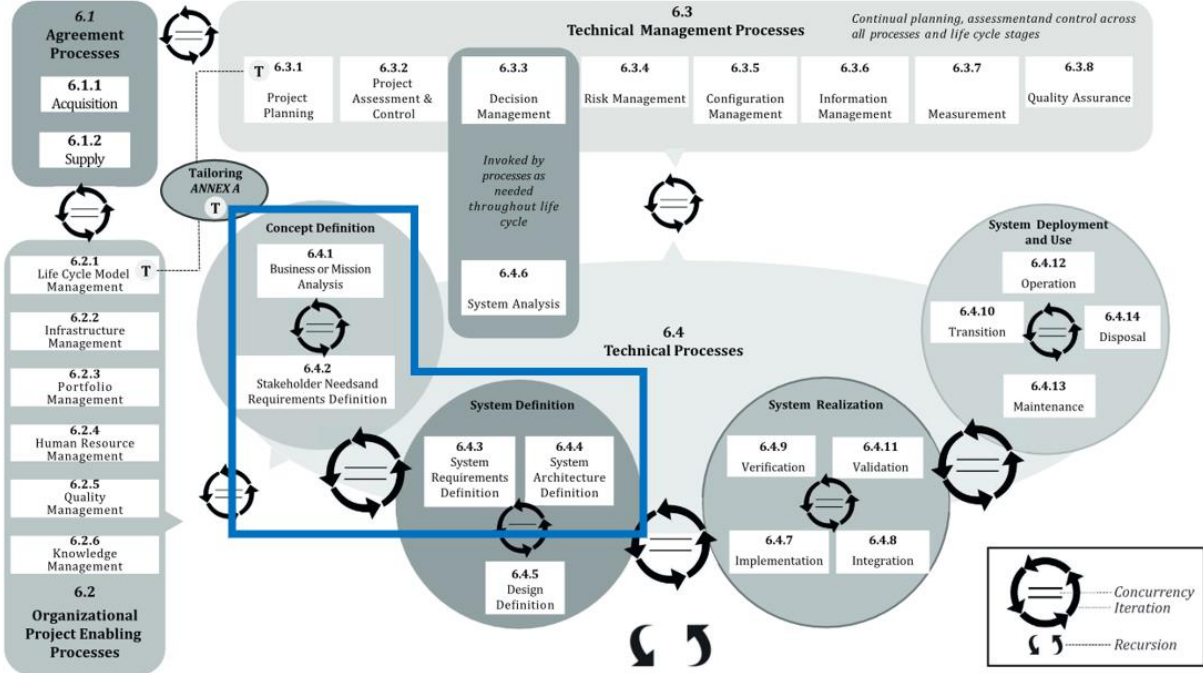


Figure 3 Interrelationships between processes (ISO/IEC/IEEE 15288 2023)

The scope of this thesis is the systems engineering technical processes, especially the mission analysis, stakeholder needs definition, system requirements definition, and system architecture definition processes (ISO/IEC/IEEE 15288 2023) (see Figure 3). Previously, we introduced MBSE as a model-based approach to systems engineering; that being said, MBSE does not stand for a subdivision, nor a subset, of systems engineering. MBSE is a complete approach to systems engineering, including systems engineering activities and considerations of different engineering disciplines. However, as far as we go till the end of this thesis, we are currently focusing on the definition of system architecture. The architect, who works for a client and with a builder, engages in a joint exploration of requirements which have yet to be stated more than vaguely. He designs satisfactory and feasible solutions to an ill-structured problem where goals and means are known with little certainty. In contrast, the engineer, who works with the architect and for a builder, seeks an optimal design solution to a clearly defined set of requirements. System architecting is creating a system architecture with the aim that the

system to be built will be ‘fit for purpose’, that is, to strive for fit, balance, and compromise among the tensions of client needs and resources, technology, and multiple stakeholder interests. Systems architecting is the front end of systems engineering (Rechtin 1991) and may be generally summed up as the art or science of constructing systems for human use. Model-Based Systems Architecting (MBSA) is often used in place of MBSE. Is MBSA a subset of MBSE, or a superset of MBSE, or are they logically completely distinct? As Systems Architecting is the front end of Systems Engineering, MBSA is the front end of MBSE. Regarding MBSA, the scope of this thesis is to build models for the systems engineering technical processes identified in Figure 4.

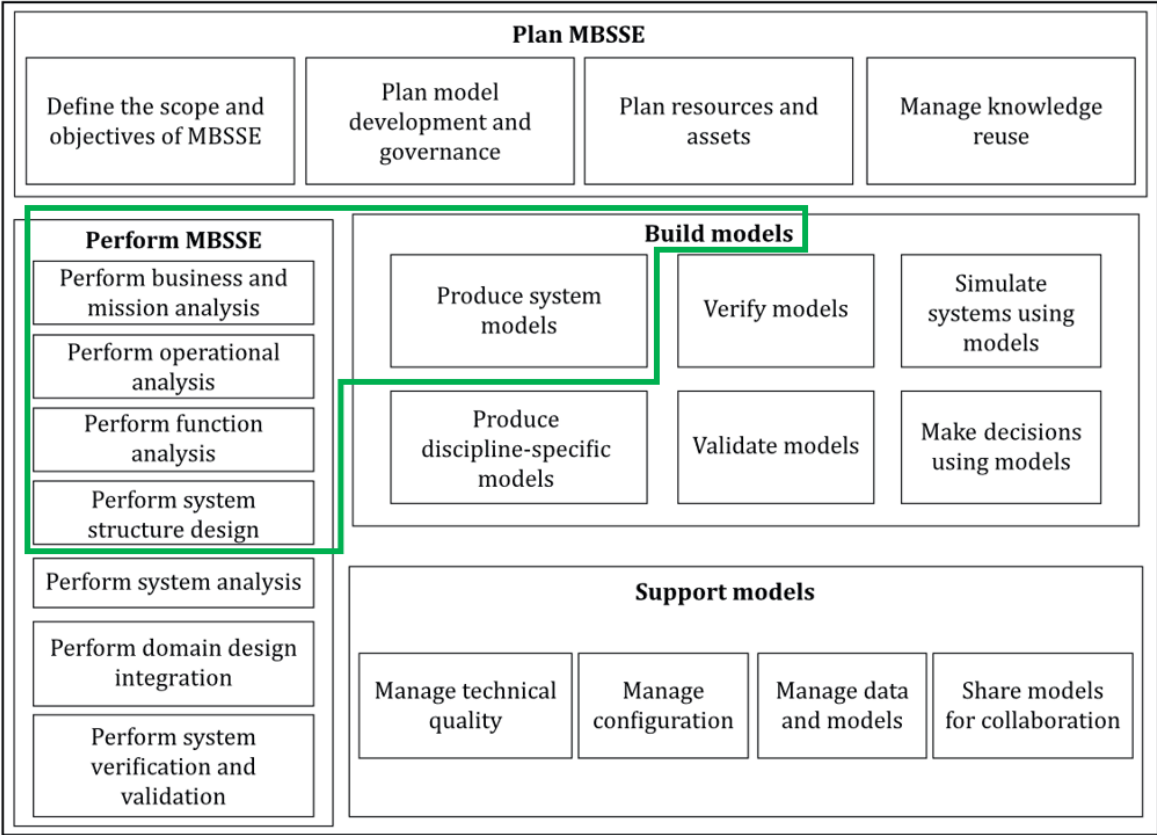


Figure 4 Model-Based Systems and Software Engineering (MBSSE) Reference model (ISO/IEC/IEEE 24641 2023)

Based on the Journey to Validation (Isaksson et al. 2020), Table 1 gives the gist of this thesis.

Table 1 Gist of the thesis

Practical context	The situation in which system architects engage in Model-Based Systems Architecting (MBSA) activities
Practical problem	Adoption challenges with the MBSA methodology for system architects
Example	The views in a complex system quickly become unintelligible for system architects.
Problem	MBSA methodology adopting challenges and lack of attention to the visual syntax of an MBSA language
Research gap	There is no virtual reality-based modelling interface for architecting engineered systems and no empirical comparison with a conventional diagrammatic modelling interface.
Research questions	Does the proposed interface have good User Satisfaction/Visual Notation/Presence/Cognitive Load/Group Dynamics?
Hypothesis	3D immersive MBSA interface will improve the performance of MBSA activities by enhancing the cognition process
Claim	A new 3D immersive MBSA interface will improve the performance of MBSA activities.
Contribution to knowledge	The new visual syntax for MBSA conceptual modelling language
Contribution to practice	How to evaluate a 3D immersive MBSA interface

MBSE, and therefore MBSA, methods and tools borrowed from software engineering are leading to various difficulties, such as:

- Increasingly complex human-model interaction forces users to invest more effort in modelling tasks than engineering activities.
- Data about a particular system is distributed across several views and models, which require learning and accessing various heterogeneous modelling languages and software.
- Data is highly codified with domain-specific languages, which makes it difficult for all stakeholders to understand.
- Substantial volume of data.
- Mainly esoteric diagrammatic representations.

This research project to improve the design of complex engineered systems relates to combining different bodies of knowledge, including systems engineering, human-computer

interaction, ergonomics, and cognitive science, with long-term fundamental research questions and objectives, such as:

- To understand how visual representations impact systems engineers' activities.
- To understand how visual representations impact systems engineers' cognition.
- To develop guidelines to specify, design and evaluate visual representations that positively impact systems engineers' activities and cognition.

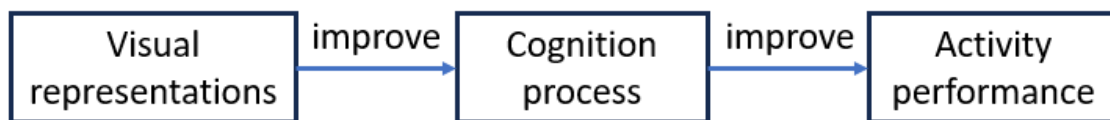


Figure 5 Overview of visual representations, cognition process, and activity performance

1.2 Research Problem

The MBSE approach was initially proposed to help people do better systems engineering activities, especially with extensive and complex systems crossing various fields. As previously discussed, it is widely applied in many famous companies. However, papers argue that the application of this approach improves complex project management (Friedenthal, Moore, and Steiner 2014; Delligatti 2013; Estefan 2007), while others say that the MBSE adoption in the real world still struggles with issues (Hein et al. 2011; Albers and Zingel 2013; Parrott et al. 2016). Those issues impede MBSE adoption from the R&D domain to the industrial domain.

First, to understand the practical problem, we need to understand the role of a system architect. A system architect defines – models, simulates and documents - each of the system's views (e.g., operational, specification, functional, logical, etc.). On the other hand, a system architect's role as orchestra conductor requires them to communicate the various system views with all corporate functions (e.g., corporate management, project management, marketing, quality, production, support, etc.). Ideally, a system architect should be able to achieve those tasks without any obstacles. Despite the increasing adoption of MBSE solutions, a significant proportion of architects continue to develop engineered systems with Microsoft Office or Google software suites with diagrams that are drawings rather than models created with a modelling language (syntax and semantics) that a processor can check. The following paragraphs will systematically start with the desired situation by a system architect before discussing the current practices and their limits.

[Mushrooming of models] A system architect expects that an MBSE approach facilitates the definition of architecture from different consistent viewpoints (requirements, operational, functional, logical, [...], structural) gathered in an architecture framework (e.g., Magic Grid, OOSEM, MOFLT, DoDAF, MODAF, NAF, etc.). Each view is a set of models that permits the

separation of concerns regarding the systems of interest. Today, for a complex system, these views quickly become unintelligible mainly because of the number of model elements and their interdependencies, which negatively affect the cognitive performance of stakeholders (Baduel et al. 2018).

[Learning modelling languages and software] A system architect wants to define the architecture views in a unique modelling environment with a limited effort to learn modelling languages and software. However, as discussed, the mushrooming of models requires the development skills in numerous modelling languages and software, which implies that practitioners need to consume extra time and effort to get familiar with these languages and software while reducing efficiency and restricting communication between stakeholders (Baduel et al. 2018; Chami and Bruel 2018).

[Heterogenous models] A system architect expects the heterogeneous MBSE data to be integrated into a centralised data hub. However, combining different modelling languages and software is still challenging. Model exchange is not standardised and practically impossible between conventional MBSE software (Chami and Bruel 2018). However, the modelling of each view with dedicated modelling software leads to a heterogeneous environment that challenges the digital continuity and the consistency of digital threads.

[Lack of execution] The system architect needs to get an overview of all the information from every point of view, which is all connected. The previous issues make managing the dependencies inside a system or between sub-systems hard. When ordering, it is easy to get lost in the countless models and connections between models. So, most models used to define and manage a system architecture are inert representations but cannot be simulated dynamically to make the management more intuitive and visual. SysML-like models are static diagrams used to communicate information. Recent research studies strive to integrate simulation capabilities within MBSE modelling environments. This refers to the interoperability barriers and challenges of weak integration between tools. For one MBSE approach-based tool, it is impossible to model the system from a different point of view and execute a simulation that dynamically demonstrates the system's performance. Still, those are more palliative pragmatic patches than solutions natively designed with a system thinking approach, enabling a system architect to simulate the preliminary design of system properties.

[Influence of software] A system architect desires the tools to hide their implementation and reuse the terminology and fundamental concepts of systems engineering. Applying the MBSE approach is appreciated as it improves the management and sharing of data architecture between stakeholders. However, systems architecture modelling languages and environments have not been invented with a “systems thinking” approach but are

opportunistically derived from existing software engineering practices (Herzog, Pandikow, and Ab 2005), where the concepts are primarily abstract and the perceived deviation on the mental model, with the tools containing functions dedicated to software engineering. For instance, SysML is a UML profile tailored for systems engineering. So, with everything that's been staged so far, we can say that the system architecture tools remain inadequate for various reasons, including, but not limited to, laborious modelling operations, rich and complex modelling language, excessive modelling freedom, poor management of multiple views of an element, overabundant models and interdependencies, etc. (Chami and Bruel 2018; Grundel and Abulawi 2016; Herzog, Pandikow, and Ab 2005).

[Lack of intelligence of modelling environments] A system architect could expect assistance from a virtual agent. Still, the lack of intelligence in complex systems architecture leads to activities lacking the "computer-aided" part. For example, even the most basic tasks, such as modelling a flow between two blocks of a SysML activity diagram, require no less than seven actions (Herzog, Hallonquist, and Naeser 2012) in some MBSE software. There is no constraint for the incompatible connection against the physical laws between the two ports. Those mistakes can occasionally consume considerable effort and time to be fixed. The "computer-aided" assistance is limited to neat diagrams but not too much in the specification, functional analysis, and partitioning of blocks. In addition, domain knowledge, such as architecture rules and best practices, remains tacit knowledge in the head of experts or MBSE frameworks highly customised by a team of experts in a company to incorporate their specific domain specificities. In contrast, the modelling environments could continuously learn to augment human intelligence, automate modelling and design routines, and recommend deep knowledge.

[Poor visual syntax] As a system architect, I want to manipulate rich visuals to holistically integrate conceptual information and communicate with all internal and external stakeholders. However, the system architect deals with tons of diagrams from different viewpoints. While the previous issues weaken the connection of the models, the visual syntax of the modelling language itself is yet another main issue that needs to be discussed. The semantic aspect of MBSE graphic notations receives much attention, whereas choosing an appropriate form comes down to personal taste. However, visual syntax has an essential influence on cognitive effectiveness, as does semantic content.

To sum up, there are still many things that could be improved about MBSE's efficiency. A system architect doesn't have a toolbox specifically designed to support its tasks, and the highly codified MBSE notations require significant effort to learn and master due to their high level of abstraction, which makes them esoteric for most people involved in the development of engineered systems.

1.3 Research Questions and Hypotheses

With the current challenges, this thesis explores and seeks to validate a central research claim that is formulated as follows:

Virtual Reality (VR) technology can support systems architecting activities by offering the opportunity:

- *to encode domain knowledge in rich 3D representations that require fewer decoding efforts for notational nonexperts in conceptual modelling,*
- *to combine all architecture viewpoints in an immersive, unbounded 3D virtual world, which removes the constraint of navigating between multiple modelling languages, software, and displays.*

In response to the need for new human-model interaction for model-based systems architecting, the broad research question established for the study was:

What is the potential of using VR technology to co-design system architectures, compared with conventional model-based systems architecting notations, software, and devices?

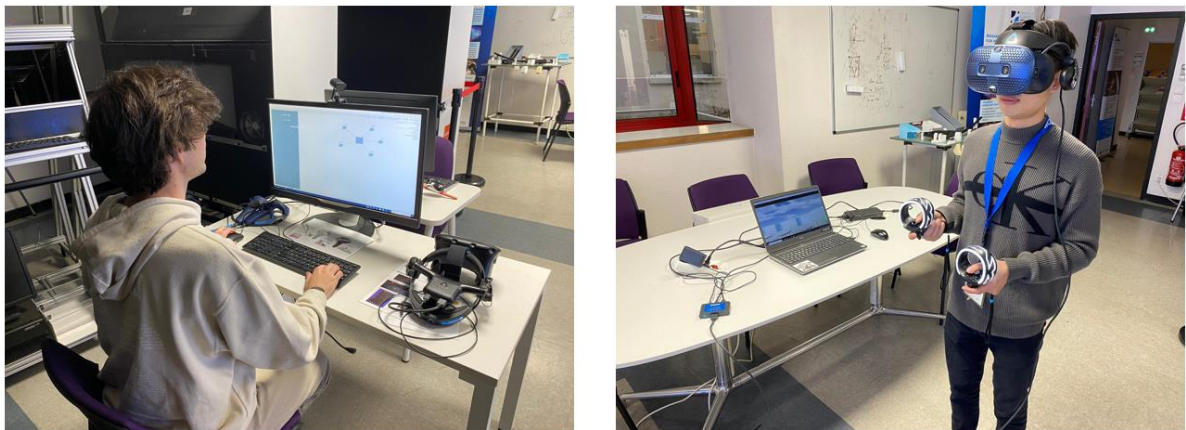


Figure 6 Conventional modelling scenario and VR technology-based modelling scenario

However, although the research question is relevant, it needs to be more focused and testable. It needs to be narrower and more specific to serve as a research gap where specific knowledge contribution claims can be made. Thus, the previous generic research question leads to the statement of several better-defined research questions with the related Null (H_0) and alternative (H_1) hypotheses. The research questions are grouped into categories corresponding to high-level evaluation metrics: user satisfaction, quality of the visual notation, presence, cognitive load, and group dynamics. Those categories gradually evolved as the research progressed.

In this thesis, the proposed solution will be the “3D immersive MBSA interface”, including a newly proposed stereoscopic visual syntax from which end-users interact using a VR headset and controllers. The control group will use a conventional “2D diagrammatic MBSA interface”, including a monoscopic visual syntax from which end-users interact using mouse and keyboards.

- **User satisfaction**

Yigitbas et al. (2022) uses user satisfaction to evaluate usability. Additionally, this paper has an extra questionnaire covering other aspects (e.g., usage, fun, and motivation) to receive general feedback and remarks. This also aligns with the idea of user satisfaction assessment. Therefore, in this thesis, user satisfaction covers usability and other complementary aspects: perceived efficiency, perceived effectiveness, perceived retention rate, intention to use in the future, usage fun, and user motivation.

RQ1: Does the 3D immersive MBSA interface increase user satisfaction for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface?

H₀: The 3D immersive MBSA interface doesn't increase user satisfaction for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface.

H₁: The 3D immersive MBSA interface increases user satisfaction for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface.

- **Quality of the visual notation**

HCI combines multiple elements, including visualisation and interaction devices, semantic constructs (concepts), a visual notation, [...], and a modelling process. Whereas user satisfaction relates to the evaluation of the whole solution, we want to investigate the influence of the quality of the visual notation by asking the research question:

RQ2: Does the 3D immersive MBSA interface provide better visual notations regarding the Physics of Notation (PoN) (Moody 2009) than the 2D diagrammatic MBSA interface?

H₀: The 3D immersive MBSA interface doesn't provide better visual notation regarding the PoN than the 2D diagrammatic MBSA interface.

H₁: The 3D immersive MBSA interface provides better visual notation regarding the PoN than the 2D diagrammatic MBSA interface.

In addition to evaluating the visual notations with PoN, we also want to know if the expertise level of the participants in MBSA/SA/VR will affect their assessment by asking another research question:

RQ2.1: Does the expertise level in MBSA/SE/SA/VR influence the evaluation of visual notation?

H₀: The expertise level in MBSE/SE/SA/VR doesn't influence the evaluation of visual notation.

H₁: The expertise level in MBSE/SE/SA/VR influences the evaluation of visual notation.

The 3D immersive MBSA interface with the 3D representations may enhance the understanding of spatial relations. Therefore, we added one more research question related to the spatial relation:

RQ2.2: Does the extra information about the spatial relation provided by the 3D immersive interface improve the MBSA activities?

H₀: The extra information about spatial relations provided by the 3D immersive interface does not improve MBSA activities.

H₁: There is an improvement in MBSA activities with the extra information about spatial relations provided by the 3D immersive interface.

- **Presence**

One of the essential characteristics of virtual reality is creating a sense of presence. Immersion refers to the objective level of sensory fidelity a virtual environment provides, while presence refers to a user's subjective psychological response to a virtual environment (Slater and Wilbur 1997). Thus, presence or psychological realism is a psychological, perceptual, and cognitive consequence of immersion. Presence is the psychological perception of "being in" or "existing in" the virtual environment in which one is immersed. In most cases, the improved sense of presence will lead to better task performance (Lee 2004; Slater et al. 1999).

RQ3: Does the 3D immersive MBSA interface possess a sufficient level of presence so that the users can stay focused on their tasks while in the virtual world?

H₀: The 3D immersive MBSA interface lacks sufficient presence so that users can stay focused on their tasks in the virtual world.

H₁: The 3D immersive MBSA interface possesses a sufficient level of presence so that the users can stay focused on their tasks while in the virtual world.

- **Cognitive load**

One change with the transformation from a 2D diagrammatic MBSA interface to a 3D immersive MBSA interface is to improve the visual syntax. A better visual syntax will help in diagrammatic reasoning, which means it will reduce the cognitive load. However, more than one factor will impact the cognitive load. While in the exploring phase with the 3D MBSA

interface, we are currently focusing on the visual syntax side, with the other factors remaining mysterious to discover in the future. We want to evaluate the cognitive load to find out if this 3D immersive interface can reduce the cognitive load and improve the MBSA activities performance:

RQ4: Does the 3D immersive MBSA interface reduces the cognitive load for co-designing a system architecture compared with a 2D diagrammatic MBSA interface?

H₀: The 3D immersive MBSA interface doesn't reduce the cognitive load for co-designing a system architecture compared with the 2D diagrammatic MBSA interface.

H₁: The 3D immersive MBSA interface reduces the cognitive load for co-designing a system architecture compared with the 2D diagrammatic MBSA interface.

- **Group dynamics**

The word “cooperation” is frequently used in the MBSE activity context. Indeed, given MBSE's interdisciplinary nature, communication among diverse teams with expertise in different disciplines is inevitable. Based on (Darses et al. 2001), the forms of collaboration can be categorised into two types: distributed design situations and co-design situations. Distributed design situations refer to situations where “the actors of the design are simultaneously, but not together, involved in the same collective process”. Co-design situations refer to situations where “*the design partners develop the solution together: they share an identical goal and contribute to reach it through their specific competencies*”. In this thesis, the co-design situations are focused. Cooperation promotes a holistic understanding of complex systems, aligns requirements across disciplines, ensures model consistency, and more. It is critical for successfully managing complex MBSE projects, so we want to evaluate it with the proposed 3D immersive HCI.

RQ5: Does the 3D immersive MBSA interface increase the group dynamics for co-designing system architecture activities compared with a 2D diagrammatic MBSA interface?

H₀: The 3D immersive MBSA interface doesn't increase the group dynamics for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface.

H₁: The 3D immersive MBSA interface increases the group dynamics for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface.

Although these research questions are more precise, analysing the potential of using VR technology for architecting systems based on lower-level criteria such as usability, user satisfaction, quality of the visual notation, effectiveness, usefulness, [...], the cognitive load remains very broad for providing objective empirical evidence. Indeed, it becomes difficult to conduct a fair comparison with existing practices, especially when the number of modalities

(stereoscopy/monoscopic, visualisation device, interaction device, visual metaphors, interaction metaphors, conceptual modelling language, diversity of systems architecting tasks, level of expertise in all disciplines, etc.) is tremendously large for controlled experiments. In this thesis, we had two opportunities related to the trade-off between the accuracy of the answers to the research questions and the breadth of the questions. First, we could extensively reduce the scope to a single criterion (e.g., effectiveness) and two modalities (e.g., comparing SysML semantics and visual syntax in a monoscopic desktop-based environment vs. SysML semantics and visual syntax in a stereoscopic head-mounted display). But this choice brings a lot of questions, such as: what is the value of conserving the same visual metaphors (blocks and arrows) in an immersive environment? The second option was to conduct exploratory research by assessing various dimensions that may influence the potential of using VR for systems architecting by answering weaker research questions. By relaxing the test environment and experimental procedures, we could better mimic real-world MBSA situations (increase of external validity or breadth of questions) while increasing the number of uncontrolled sources of variation (reducing internal validity or accuracy of answers). Ultimately, we adopted the second option to help us better understand the research context and develop more precise research questions and methods for future work.

1.4 Research Method

After the initial presentation of the generic research question, a literature review was conducted on MBSE, virtual reality, and the visual syntax of conceptual modelling notations. The primary aim was to understand the existing research and debates relevant to these particular topics and to explore studies situated at the intersection to offer insights into the application of MBSE activities in an immersive environment.

Additionally, relevant papers were explored to understand potential evaluation methods and to select validation criteria, which helped us refine the generic research question into five testable research questions. The next step is establishing a validation method according to those five questions. The plan is to conduct a vertical result analysis for each type of experiment, followed by a horizontal summary of the results of all experiments. Figure 7 illustrates the research method.

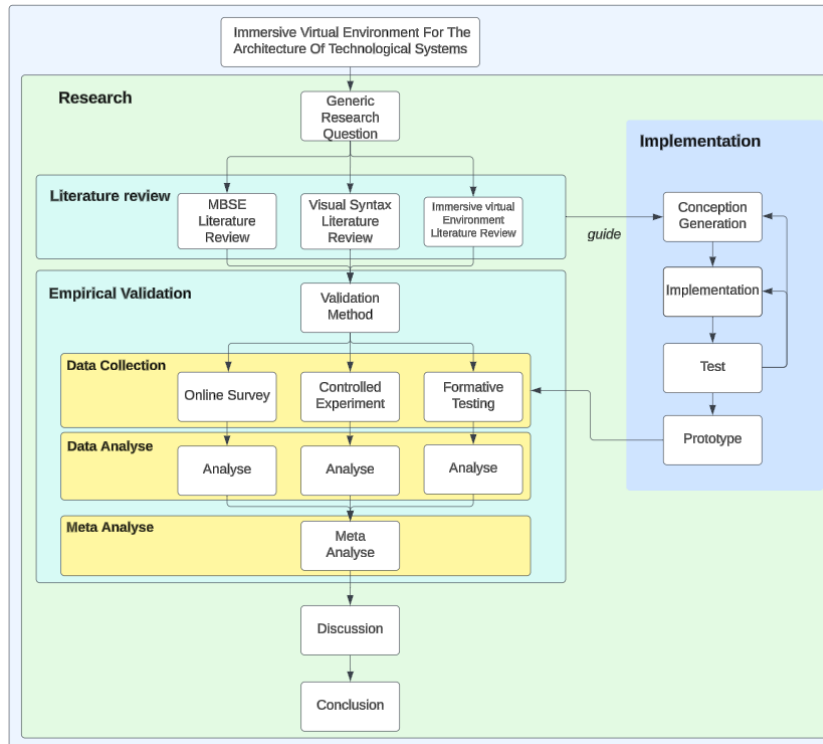


Figure 7 Overview of the research method

1.5 Statement of Contribution

As exploratory research, the scope of this thesis includes many domains. We embraced various variables to get a preview of what it will be like to have a more thorough integration of VR and the MBSE approach. As a result, as discussed in section 1.3, the scientific grounds may need to be revised. Still, we can provide a general orientation for future research where we can reduce the variable to have a more solid result. The contributions of this thesis are listed as follows:

- **Concrete integration of MBSA and VR in a unique framework**

The originality of this research is the tight integration of MBSA and virtual reality. It is more than an attempt to present multiple 2D diagrams with different layers situated in different depths on a 2D screen, to transform the visualisation of lines and boxes from 2D to 3D, or to review projects in 3D space from a particular perspective of view without the modelling capabilities. The solution is an authoring framework for systems architecting encompassing a systems engineering ontology and method, a modelling method and a fully functional software prototype.

- **New visual syntax**

The proposed immersive environment harnesses the advantages of a 3D environment, taking inspiration from the conventional MBSA ontologies, languages, methods, and software.

A new visual syntax is presented within this immersive environment, incorporating common semantic constructs derived from popular MBSE methodologies. Suppose we group MBSE into three main areas, as illustrated by Holt (see Figure 8). In that case, the thesis contributes to the third group, visualisation, wherein we present information using a new graphical notation using 3D graphics (informationally equal to the “diagrams”) to enable non-experts (e.g., management, marketing, quality, operations, safety, etc.) to contribute to the development of a system architecture actively. This new visual notation retains the universal constructs of existing MBSE methodologies but represents an evolution in visual syntax. Diagram no longer refers to the conventional 2D canvas where the engineer can create visual representations of the engineering concepts forming part of the model. Instead, it extends to an additional dimension, transforming into a virtual graphical scene that engineers can build using the new proposed 3D visual notation.

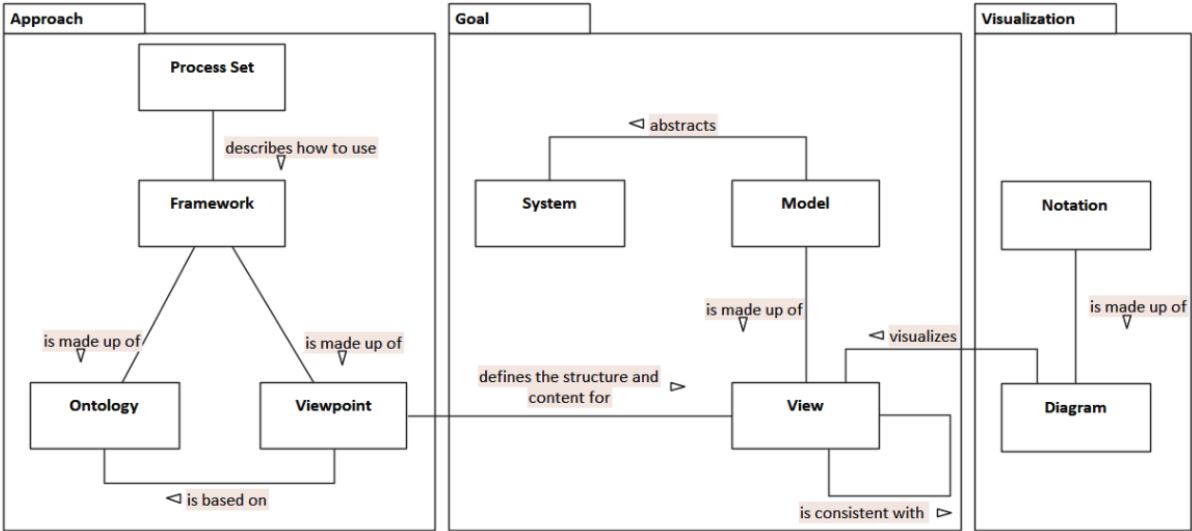


Figure 8 MBSE in a slide (Holt 2021)

- **New modelling interface in VR**

Moreover, supporting the new visual syntax necessitates the development of modelling software that can help it. Thus, implementing a new tool with an immersive environment and interface is another contribution beyond research (see Figure 9). The new immersive environment will stage the newly proposed 3D visual that encodes MBSE concepts. The interface is designed to enable interaction between the user and the environment.

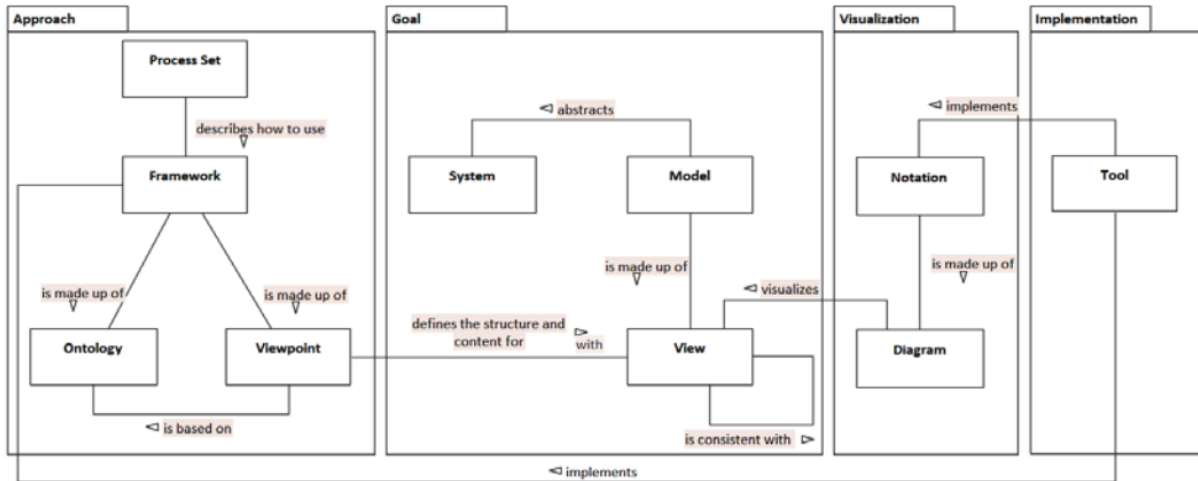


Figure 9 Expanding "MBSE in a slide" to include implementation (Holt 2021)

- **Evaluation of the 3D immersive MBSA interface**

3D immersive MBSA interfaces barely exist in the current studies. The existing attempts to develop functional prototypes of new human-computer interaction show a dramatic lack of validation. In this study, we spent a significant amount of the research effort reviewing and selecting criteria that can reflect the performance of the MBSA immersive virtual environment from the current studies in the MBSA, conceptual modelling and virtual reality areas.

In addition, various experiments conducted with the prototype contributed to evaluating the new 3D immersive MBSA interface with different complementary methods that enabled us to perform a meta-analysis to provide convincing evidence and identify conflicting claims. For example, the online survey reveals the visual differences between the new 3D graphic visual syntax and the conventional diagrammatic visual syntax using PoN principles. The controlled experiments explore the general performance between the 3D immersive MBSA interface and the 2D diagrammatic MBSA interface. The formative study gathers the experienced industrial practitioners' feedback on the 3D immersive MBSA interface. The meta-analysis across different methods comprehensively evaluates the 3D immersive MBSA interface.

- **Guidance for future work related to the 3D immersive MBSA interface**

This thesis aims to offer insights into how effectively the chosen validation criteria can reflect the performance of an MBSA activity-oriented immersive virtual environment. This will be followed by a discussion addressing the five specific research questions and a broader conclusion encompassing the research results. The research results of this thesis can contribute to developing a theory that unifies the methodical design of virtual environments with the impact of visual and interaction metaphors on cognitive (especially perceptual) human processes and the performance of system architect's tasks.

1.6 Outline

After this introduction, section 1 will present a literature review covering general conceptual modelling, Model-Based Systems Engineering (MBSE), and Human-Computer Interaction (HCI), including Virtual Reality. Section 3 will present the new interface, beginning with the main tasks of a system architect that the new interactive interface aims to support, and then the details of the software prototype, including the data model and the design of visual and interaction metaphors. Section 4 will list the selected criteria and review the various validation methods, including an online survey, two controlled experiments, a questionnaire, and formative studies. A meta-analysis will also combine results obtained with the different validation methods. Section 5 will table some reflexive discussions based on the results presented in Section 4 and some limits. Finally, Section 5 will conclude the thesis and provide perspectives for future works based on the vision of future MBSA interfaces and situations.

2 LITERATURE REVIEW

This section will present the literature review on conceptual modelling, model-based systems architecting, immersive human-computer interaction, and validation of immersive human-computer interaction, with each domain followed by a subsection for discussing the current status of the bodies of knowledge concerning the research questions.

2.1 Conceptual Modelling

2.1.1 Introduction to Conceptual Modelling

Conceptual modelling is essential in information systems, software engineering, and data design. It is the foundation for designing and developing complex systems by visually representing the system properties and its context. Conceptual modelling has three fundamental dimensions (Thalheim 2009):

- **Modelling language constructs** are applied during conceptual modelling. Their syntactics, semantics, and pragmatics must be well understood.
- **Applications domain gathering** helps to understand the problems to be solved, the opportunities of solutions for a system, and the requirements and architecture that might be prescribed for the solution that has been chosen.
- **Engineering** is oriented towards encapsulating experiences with design problems pared down to a manageable scale.

It is essential to understand the purpose and scope of the model and the intended audience since the conceptual model is differently defined in different domains, such as computer science and engineering. There has yet to be a common notation of the conception of a conceptual model. The same is valid for understanding the idea of the model (Thalheim 2018). Different notations define the models from a similar perspective in other domains.

Form-Function: Function is defined as “intended behaviour” in (Gero and Kannengiesser 2004), and form is mentioned as “The essential mode of reasoning in designing is to reason from function to form” in (Roozenburg and Eekels 1995)

Form-fit-function (3F) (Dumas and Mintzberg 1991) is a manufacturing terminology used to describe a part's characteristics. It addresses the debate over form versus function in design, where form stands for shape and materials, fit stands for connector and interface, and function stands for delivering the designed performance.

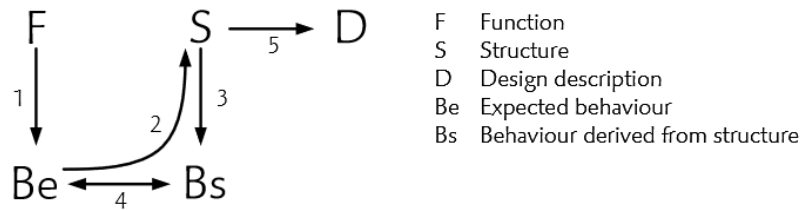


Figure 10 The FBS framework after (Gero and Kannengiesser 2004). Processes: (1) formulation, (2) synthesis, (3) analysis, (4) evaluation, (5) documentation. Reformulation processes, which are included in the original version of this figure, have been omitted for clarity (Müller 2020)

Function-Behaviour-Structure (FBS) is the framework proposed by (Gero and Kannengiesser 2004). “When a product interacts with its environment, i.e., when it is used, it shows a behaviour” (Müller 2020). The expected behaviour is denoted “Be”. Evaluating the match between Bs (from “structural behaviour”, i.e., actual behaviour) and Be requires form analysis. Figure 10 illustrates the relations between function, behaviour, and structure.

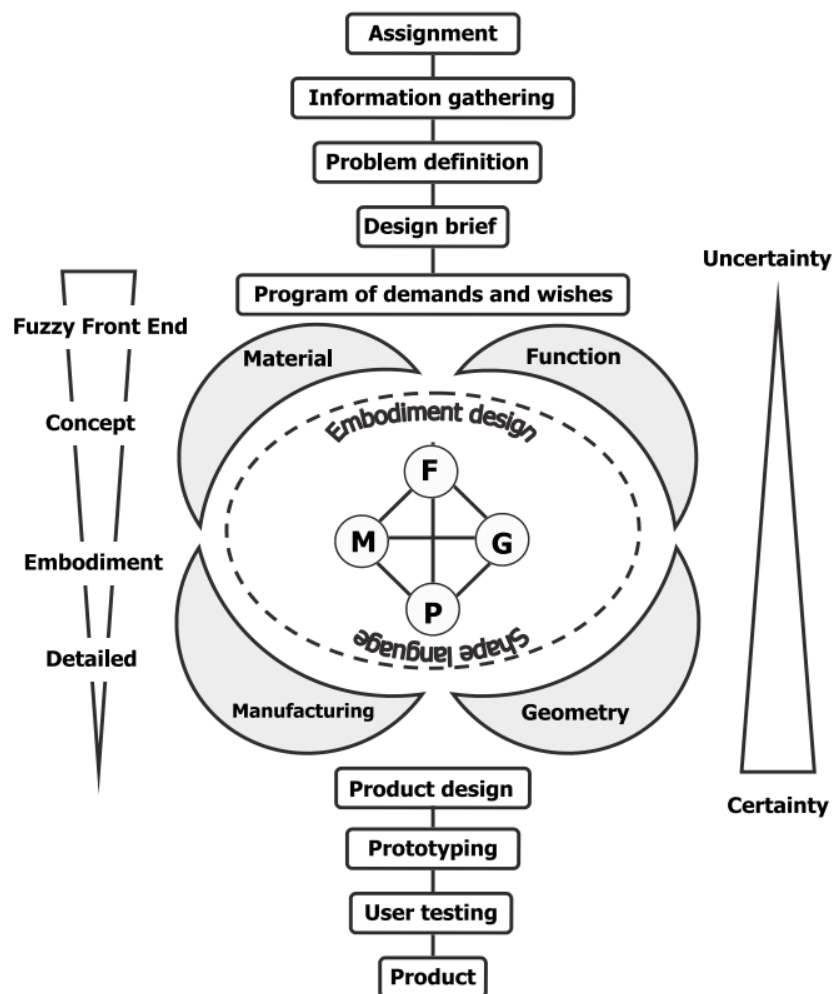


Figure 11 Design process with embodiment design (Langeveld 2011)

Embodiment design refers to the product design and development phase, where abstract concepts and ideas are translated into physical form or structure. It involves defining a product's physical characteristics, specifications, and overall architecture based on the conceptual design. Due to the overlaps of the activities in conceptual design, it is hard to separate the conceptual design and embodiment design. Pugh (Pugh 1991) merged embodiment design into the conceptual design. Embodiment design can cover product architecture, configuration design, and parametric design (Sapuan 2017). Figure 11 illustrates the embodiment design (Langeveld 2011).

Physical systems and software differ because software modelling doesn't include the physical design process. The modelling process with the software domain is purely abstract, while systems engineering involves a concrete structural design process. The CDFA method (Formentini et al. 2022) for aircraft system architectures comprises the product functional decomposition and architecture geometrical definition (see Figure 12). The proposed 3D immersive MBSA interface includes the physical design as it aims to support the preliminary modelling process of technological systems.

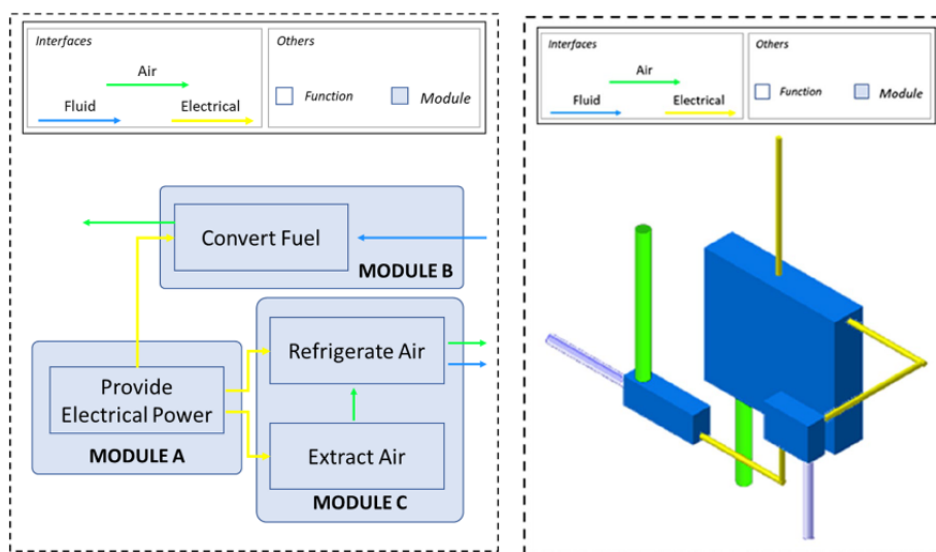


Figure 12 Product functional decomposition allocated to logical subsystems (left) and architecture geometrical definition (right)

System architecture is a system's fundamental concepts or properties in its environment embodied in its elements, relationships and the principles of its design and evolution (ISO/IEC/IEEE 42010 2011).

Another perspective of conceptual modelling is the collective aspect. Various presentations representing complex systems from different viewpoints aim to facilitate communication between stakeholders involved in the design process. This leads to the conception of collective cognition, which is defined by (Darses and Falzon 1996) as “*the intensive implementation of*

representation transfer functions, which enable the sharing of representation transformations in the collective problem space, and the integration of these multiple representations, not only in the representation phase of the problem but also in the description of the solution, which can also be seen from different angles”.

Summary of Conceptual Modelling

Conceptual design, embodiment design, and system architecture are often synonyms. Embodiment design is more prevalent in the engineering design community, whereas system architecture is the preferred term in the systems engineering community. Both pursue the same objectives. However, by their origins, embodiment design is more often exemplified on mechanical or mechatronic products, which are relatively simple from a systemic perspective. System architecture is usually broader than embodiment design as it pays more attention to the big picture by analysing the system and its context from different architectural viewpoints following a top-down approach. Most system architecture descriptions are also more abstract than embodiment design representations. They can be used to describe all sorts of systems and systems of systems with a higher level of abstraction before starting the definition of preliminary shapes and interfaces. System architecture design is also standardised (ISO/IEC/IEEE 42010 2022; ISO/IEC/IEEE 42020 2019; ISO/IEC/IEEE 42030 2019) and relies on fundamentals of the systems thinking principles (Rick et al. 2023) (see Appendix A). In contrast, embodiment design is more of a set of academic practices.

2.1.2 Graphical Conceptual Modelling Notations

A visual notation (or visual language, graphical notation, diagramming notation) consists of a set of graphic symbols (visual vocabulary), a set of compositional rules (visual grammar), and definitions of the meaning of each symbol (visual semantics). The visual vocabulary and grammar form the visual (or concrete) syntax. Graphic symbols symbolise (perceptually represent) semantic constructs, typically defined by a metamodel (“ISO/IEC, ISO/IEC Standard 24744: Software Engineering—Metamodel for Development Methodologies”, 2007). The meaning of graphical symbols is determined by mapping them to the constructs they represent. A valid expression in a visual notation is called a visual sentence or diagram. Diagrams are composed of symbol instances (tokens) arranged according to the rules of visual grammar (Moody 2009).

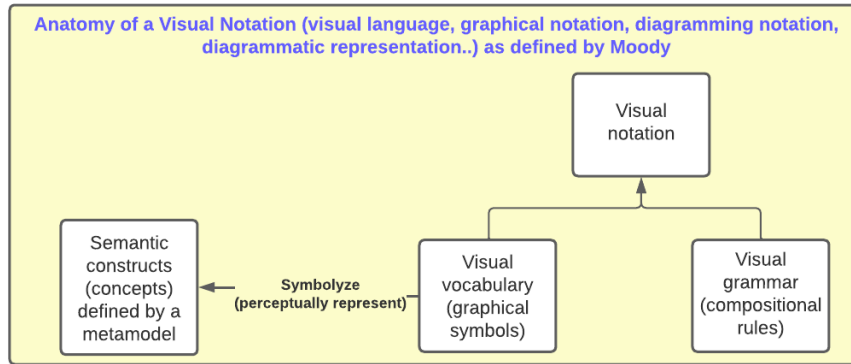


Figure 13 Illustration of Visual notation

A visual notation differs from the textual language. Textual language uses sequences of characters to encode information linearly. In contrast, visual languages rely on the spatial arrangement of graphic and textual elements to convey information in a two-dimensional format (Moody 2009). A former definition can be found in (Larkin and Simon 1987):

- A data structure in which elements appear in a single sequence is what we will call a sentential representation.
- A data structure in which a two-dimensional location indexes information is called a diagrammatic representation.

Visual graphics are essential in visual notation but should be discussed (Siau and Cao 2001; Opdahl and Henderson-Sellers 2002). The graphics vary across visual notations at different classification levels, such as iconic/symbolic or 2D/3D.

The terms “iconic” or “symbolic” relate to the abstraction level of the visual graphics. Figure 14 (Grundel and Abulawi 2016) illustrates various levels of abstraction in representing a tree, serving as an example of a complex multifunctional object. The visual graphic is symbolic when it has a high level of abstraction, ranging from figurative meanings to functional descriptions. Iconic visual graphics are less abstract, ranging from figurative representations and two-dimensional sketches to virtual and physical three-dimensional models. The visual graphics with the lowest degree of abstraction are closest to the real object.

Symbolic				Iconic						Real		
← Degree of abstraction												
Paradise, Sin, Tree of life,	Liberation, Age, Spring - Youth	Genealogy Organigram	Water evaporation, Building material, Photosynthesis									
Figurative meaning (transcendent, ideologic, ulterior)	Emotions, Sensations, Associations	Figurative meaning	Functions	Imagination of the object (pictorial thinking)	Pictogram	Sketch	Watercolor, Oil-painting Scale, transmutation	Photograph, Scale	Stereoscopy 3D-CAD-Image, Scale	Hologram (scaled down), Scale	Sculpture (scaled down), Scale	Real object, Item, physical Shape
					2D, minimizing redundancies	2D, reducing redundancies	2D, divergence from original	2D, steric feel	3D-Image (virtual, real)	3 Dimensions (virtual)	3 Dimensions (physical)	Reality

Figure 14 Representation of a tree at various levels of abstraction (Grundel and Abulawi 2016)

“Symbolic” / “Iconic” conceptions derive from another term “semiotic”. One of the broadest definitions of “semiotics” (Eco 1976) states that semiotics concerns everything that can be taken as a sign. Another short version of the definition is the study of signs. Two dominant contemporary models of what constitutes a sign are those of the Swiss linguist Ferdinand de Saussure and the American philosopher Charles Sanders Peirce (Chandler 2007).

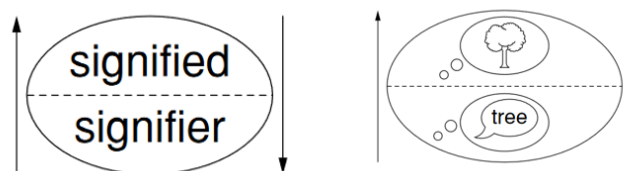


Figure 15 Saussure's model of the sign (Chandler 2007)

Saussure’s sign model is a dyadic model. He defined a sign as being composed of a “signifier” and a “signified” (see Figure 15). The signifier is the form that the sign takes, and the signified is the concept to which it refers. Peirce proposed a sign model (see Figure 16), which is a triadic model consisting of the following:

- The representamen: the form that the sign takes.
- An interpretant: not an interpreter but rather the sense made of the sign.
- An object: something beyond the sign it refers to (a referent).

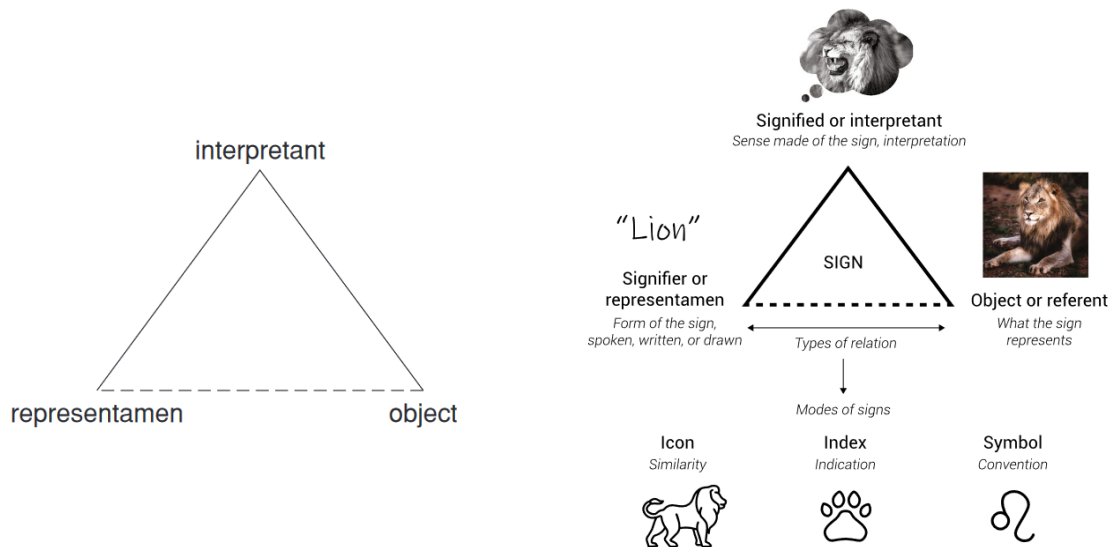


Figure 16 Peirce's semiotic triangle (Chandler 2007) on the left and sign elements and modes in semiotics (Kuhar and Polančič 2021) on the right

Meanwhile, Peirce offered a typology of signs referred to as the relationships between a representamen and its object or its interpretant him. Chandler explained it with Saussurean terms signifier and signified:

- **Symbol/symbolic:** “a mode in which the signifier does not resemble the signified but which is fundamentally arbitrary or purely conventional—so that this relationship must be agreed upon and learned: e.g., language in general (plus specific languages, alphabetical letters, punctuation marks, words, phrases and sentences, numbers, Morse code, traffic lights, national flags).”
- **Icon/iconic:** “a mode in which the signifier is perceived as resembling or imitating the signified (recognisably looking, sounding, feeling, tasting, or smelling like it) – being similar in possessing some of its qualities: e.g. a portrait, a cartoon, a scale-model, onomatopoeia, metaphors, realistic sounds in ‘program music’, sound effects in radio drama, a dubbed film soundtrack, imitative gestures”.
- **Index/indexical:** “a mode in which the signifier is not arbitrary but is directly connected in some way (physically or causally) to the signified (regardless of intention) – this link can be observed or inferred: e.g. ‘natural signs’ (smoke, thunder, footprints, echoes, non-synthetic odours and flavours), medical symptoms (pain, a rash, pulse-rate), measuring instruments (weathercock, thermometer, clock, spirit-level), ‘signals’ (a knock on a door, a phone ringing), pointers (a pointing ‘index’ finger, a directional signpost), recordings (a photograph, a film, video or television shot, an audio recorded voice), personal ‘trademarks’ (handwriting, catchphrases)”.

For the state-of-the-art conceptual modelling notations – e.g., Unified Modelling Language (UML), Entity-Relationship Diagrams (ERD), Business Process Model and Notation (BPMN), Systems Modelling Language (SysML) – the visual graphics are usually symbolic which is not well received by notational non-experts. Some studies investigate the impact of iconic visual graphics on the subject’s understanding. A study (Masri, Parker, and Gemino 2008) examines the effect of iconic visual graphics in Entity-Relationship Diagrams (Figure 17) on understanding. The results indicate that iconic graphics have a significant positive impact on understanding.

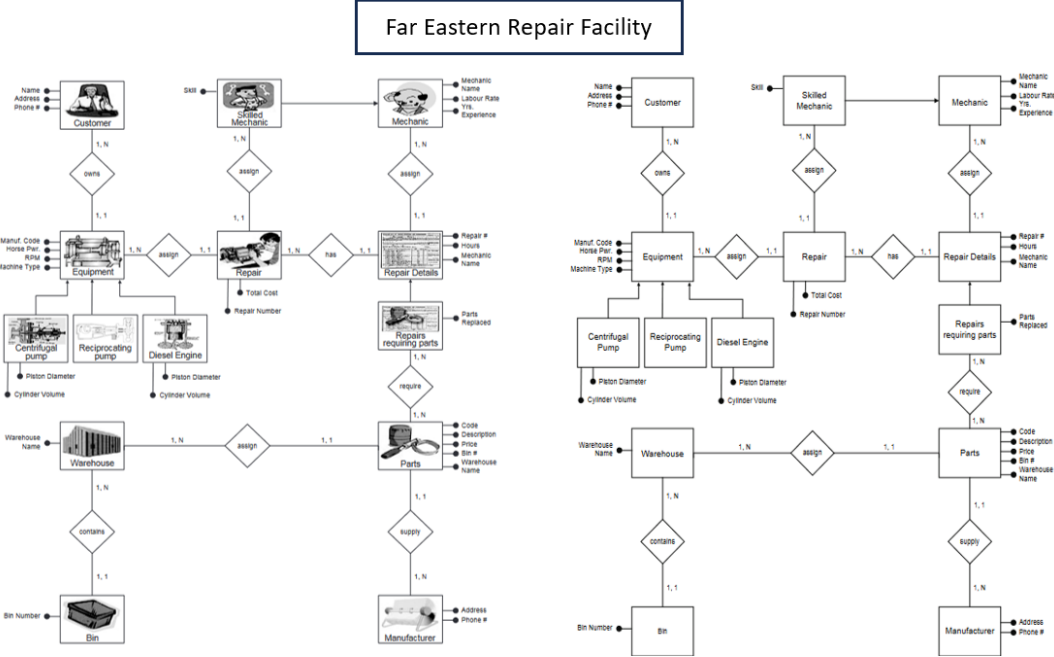


Figure 17 Iconic ERD (left) and Standard ERD (right) (Masri, Parker, and Gemino 2008)

Another paper (Grundel and Abulawi 2016) introduces the SkiPo model for developing mechatronic and mechanical systems. SkiPo visualises and documents the designer’s ideas in a systematic development process by bringing sketches into SysML (Figure 18). A sketch consists of a few essential lines representing the rough outline of the depicted object or scene. It is possible to detail the sketch with additional information, such as geometric constraints, requirements, functions, material specifications, and joining techniques (Figure 18). The overview of the SkiPo demonstrates that it can help bridge the gap between abstract function models and 3D CAD models as qualitative test results with groups of students.

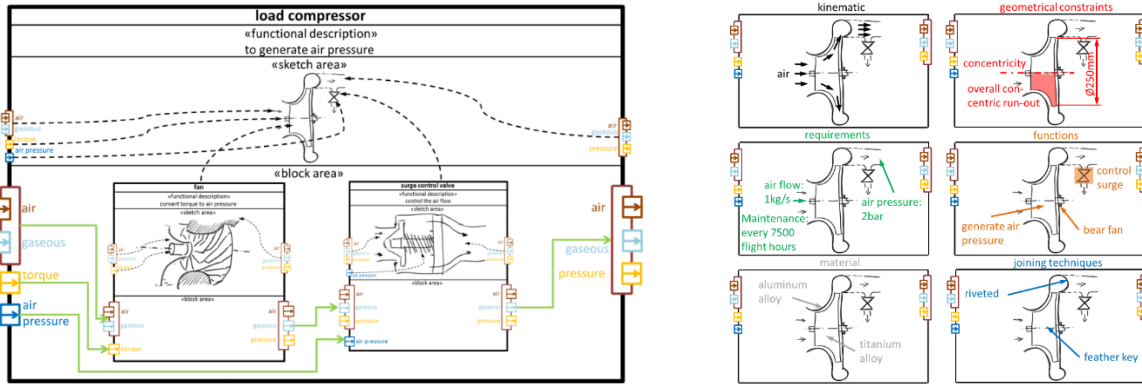


Figure 18 Example of SkiPo model and elaboration of design details (Grundel and Abulawi 2016)

In (Irani and Ware 2003), authors investigated the use of 3D shaded primitives in diagrams (as geon diagrams in this paper, see Figure 19) for better interpretation compared to 2D node-link diagrams (i.e., UML diagrams). They proposed a new substructure identification task for evaluating diagrams and used it to test the effectiveness of geon diagrams. They argued that subjects can identify the substructure shown before in the geon diagrams faster and more accurately than in the UML diagrams. It implies that using 3D primitives in diagrams makes them easier to interpret. However, what the 3D improves here is not the interpretation but only identification. Mayer defines three learning outcomes: no learning, retention (remembering), and understanding (Mayer 2001). No teaching is evident as one learning outcome. Retention refers to the capability of reproducing the presented information. Using 3D perceptual primitives helps the subjects better recall the substructures (retention). Still, it does not support the subjects at the understanding level with the low semantic transparency of the 3D perceptual primitives. The recognition and recall tasks are used to measure retention, which is the case in (Irani and Ware 2003), and transfer tests are used to measure understanding.

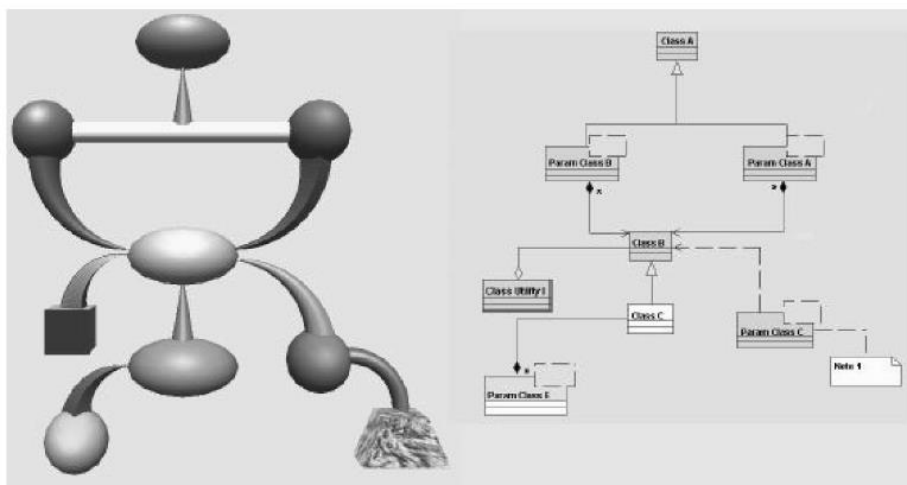


Figure 19 An example of a geon diagram and its UML equivalent (Irani and Ware 2003)

Besides the visual graphics, there are different ways of displaying graphical information: stereoscopic and monoscopic. Stereoscopic display refers to presenting left and right stereo images separately to each eye of an observer (see Figure 20). In contrast, monoscopic display only involves one single image to both eyes of an observer (Sollenberger and Milgram 1993). A paper (Hubona, Shirah, and Fout 1997) explores the difference between presenting abstract 3D data with and without stereo viewing. Results show that stereo can improve the subject's understanding of the 3D data (i.e., structural information presented by the 3D models). Studies (Martinez Escobar et al. 2015; Wai-Keung Fung et al. 2005) comparing monoscopic and stereoscopic imaging demonstrate the benefits of stereoscopic imaging with improved recognition of spatial information in the models.

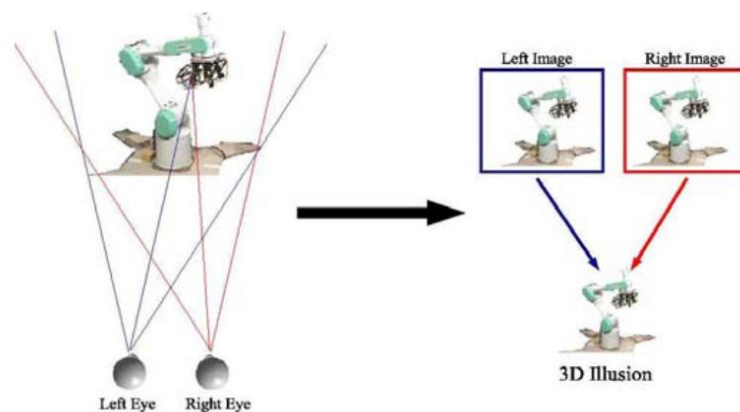


Figure 20 Illustration of stereoscopic display (Wai-keung Fung et al. 2005)

Summary of Graphical Conceptual Modelling Notations

Studies focused on improving the graphical conceptual modelling notations are presented in this section. The shared objective across these studies is to achieve efficient graphical conceptual modelling notations by modifying visual graphics. Experiments in these studies demonstrate the advantages of different visual graphics transformations (e.g., from 2D to 3D or symbolic to iconic) and diverse display approaches (monoscopic or stereoscopic). While each approach improves the graphical conceptual modelling notation, no single study introduces an approach that combines all the benefits identified in existing studies. This has sparked an interest in proposing an approach that integrates 3D stereoscopic iconic visual graphics.

2.1.3 Diagrammatic Reasoning

Diagrams are the main approach of the graphical conceptual modelling notations. Diagrammatic reasoning studies how humans use diagrammatic (pictorial) representations in problem-solving and reasoning (Chandrasekaran, Glasgow, and Narayanan 1995). A study (Larkin and Simon 1987) discussed the differences between diagrammatic and sentential

representations of information. They claim that the main difference between diagrammatic representation and sentential representation is that diagrammatic representation explicitly preserves information about the topological and geometric relations among the problem components, while sentential representation does not.

Simon (Simon 1978) proposed wholly distinct concepts of informational and computational equivalence of representations. Two representations are informationally equivalent if all the information from one representation can be inferable from another representation, and vice versa. When two representations are informationally equivalent, any inference that can be drawn easily and quickly based on the information explicitly coming from one representation can also be removed easily and quickly from another, and vice versa. These two representations are computationally equivalent. Since the ease and rapidity of inference depend on the operator's availability for modifying and augmenting data structure and the operator's speed, comparing computational equivalence between two representations must include both data and operators.

A representation consists of data structures and programs operating on them to make new inferences. The program here refers to production systems. In general, the computational efficiency of a representation depends on three factors: data structure, program, attention management, and how well they work together. Attention management is included since data structures for a problem are complex. It is crucial to provide an attention management system that determines what portion of the data structure is currently attended to and can trigger the program. The program operating on the data structure employs the following kinds of processes:

1. *Search* operates on the node-link data structures, seeking to locate sets of elements that satisfy the conditions of one or more productions. This process requires attention management.
2. *Recognition* matches the condition elements of a production to data elements located through search. Recognition depends on a match between the elements in the data structure and the conditions of the productions in the program.
3. *Inference* executes the associated action to add new (inferred) elements to the data structure.

Results (Larkin and Simon 1987) demonstrate that a diagram can be better than a verbal description when they are informational equivalence, with diagrams being more computationally efficient for solving problems. Diagrams group together information, avoiding extensive research. A diagram uses location to group information to avoid the need to match symbolic labels, and a diagram supports perceptual inferences that are easy for humans.

Furthermore, they believe the diagrams' advantages are mainly computational (i.e., any inference that can be drawn easily and quickly based on the information explicitly coming from one representation), indexing the information supporting beneficial and efficient computational processes. The advantages of the diagrams are only effective for those familiar with computational processes.

Based on this research (Larkin and Simon 1987), another study (Kim, Hahn, and Hahn 2000) explores the reasoning of multiple diagrams and where the cognitive processes in diagrammatic reasoning consist of perceptual and conceptual processes (Larkin and Simon 1987, Narayanan et al. 1995, Rogers 1996). It is similar to the process of the program operating on the data structure. The perceptual process (search and recognition), as described by (Bolles 1991), involves a bottom-up approach, where information is sensed, and its meaning and value are discerned. In contrast, the conceptual process (inference), articulated by (Simon and Lea 1974), operates from a top-down perspective, generating and refining hypotheses. In summary, the perceptual process involves searching and recognising relevant information, whereas the conceptual process involves reasoning by inference and deriving new information. They argue that the requirement for efficient perceptual and conceptual processes in reasoning with a single diagram extends to the context of multiple diagrams. When dealing with numerous diagrams, the perceptual integration process involves linking relevant items distributed across various diagrams from different perspectives. The conceptual integration process generates and refines hypotheses about the target system based on the diverse information inferred from other diagrams. They proposed two approaches, visual cues and contextual information, to improve perceptual and conceptual integration processes. Visual cues (an example in Figure 21) help to identify how an item in one diagram is related to other items in different diagrams, which decreases the mental effort to reason with multiple diagrams (Woods and Watts 1997). Contextual information (an example in Figure 23) helps to identify whether the data provided in various diagrams is essential to the attending hypothesis. One approach for contextual information is a "long shot" (see Figure 23), i.e., a big picture showing the entire system (Woods and Watts 1997).

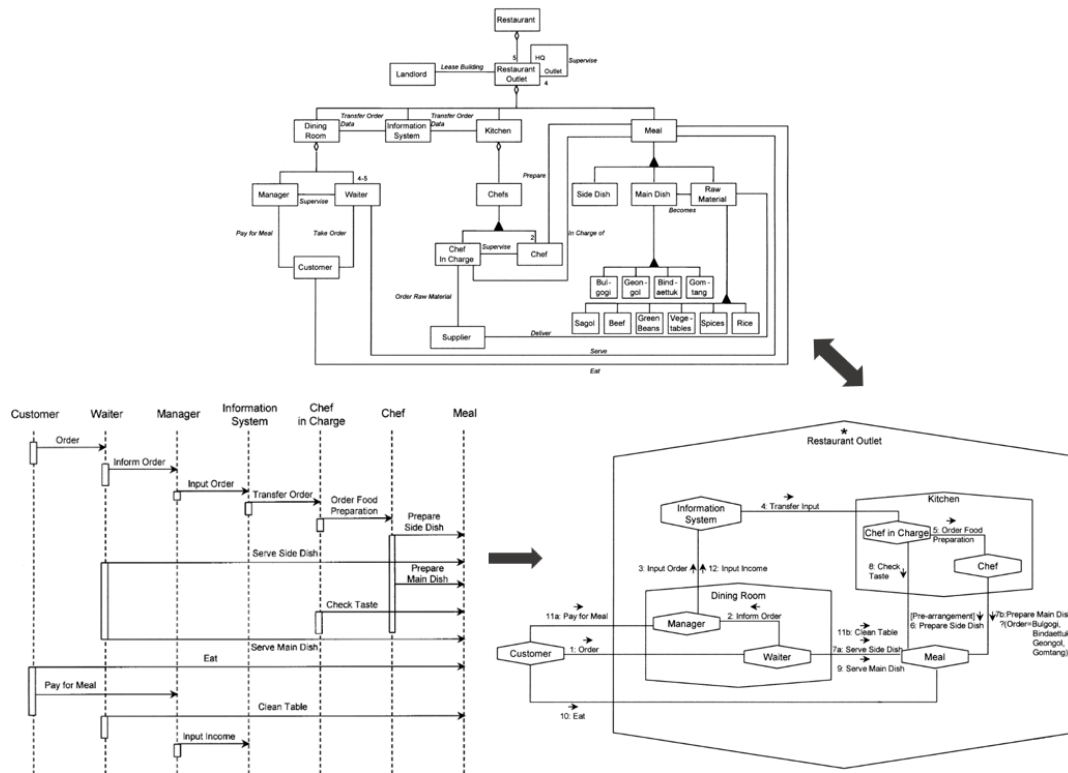


Figure 21 Modification for Visual Cues with Class Diagram (Top), Event Trace Diagram (Left), and Object Message Diagram (Right) (Kim, Hahn, and Hahn 2000)

The experiment conducted by (Kim, Hahn, and Hahn 2000) will be presented to understand better the perceptual and conceptual integration processes and the approaches as visual cues and contextual information to improve them. This experiment evaluates the effect of visual cues and contextual information on the cognitive integration process. There will be a control group and an experimental group. The control group will present the system with diagrams without visual cues or contextual information. The experimental group will present the system with some modifications to the diagrams to add visual cues and contextual information.

For the visual cues in the experimental group, the event trace diagram is replaced by an object message diagram (see Figure 21). Those two diagrams are informationally equivalent, but the object message diagram shares more visual similarities with the class diagram, which makes it easier to relate the common elements in the diagrams. A diagram transition graph (see Figure 22) is used to evaluate the perceptual cognitive activities. The transition graph describes the trajectory of the subjects' transitions among the multiple diagrams during the experiment, and this information is coded as perceptual data.

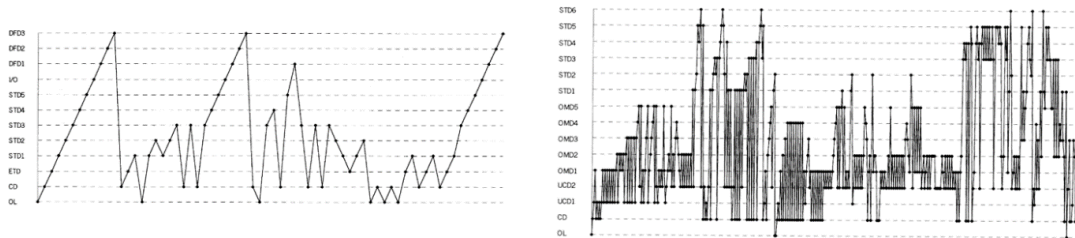


Figure 22 Diagram Transition Graph of Control Group (Left) and Experimental Group (Right) (Kim, Hahn, and Hahn 2000)

For the contextual information, a use case diagram and context diagram describe the system (see Figure 23). A context diagram is considered a “long shot” at a detailed level. It provides a broader context that allows the subject to identify the relative importance of individual items concerning the attending hypotheses. Together with the use case diagram at high-level grouping, those two diagrams help the subjects understand the detailed process diagrams with an overall view of the entire system. The hypothesis behaviour graph (Figure 24) evaluates the conceptual cognitive activities. The hypothesis behaviour graph records the subjects’ problem-solving behaviours in generating and refining the hypotheses. Each rectangle in the hypothesis behaviour graph represents a hypothesis caused by the subject, and the related diagrams are indicated under the rectangle. Another rectangle represents a refined hypothesis placed one step to the right, and the number in the bracket on the right side shows the time of refinement.

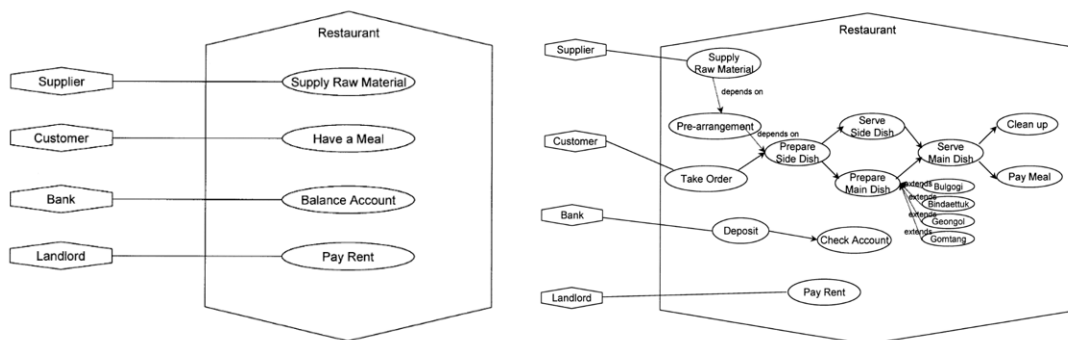


Figure 23 Modification for Contextual Information with Use Case Diagram (Left) and Context Diagram (Right) (Kim, Hahn, and Hahn 2000)

The results of this experiment show the effectiveness of improvement in understanding the target system by bringing in visual cues and contextual information for the perceptual and conceptual integration processes.

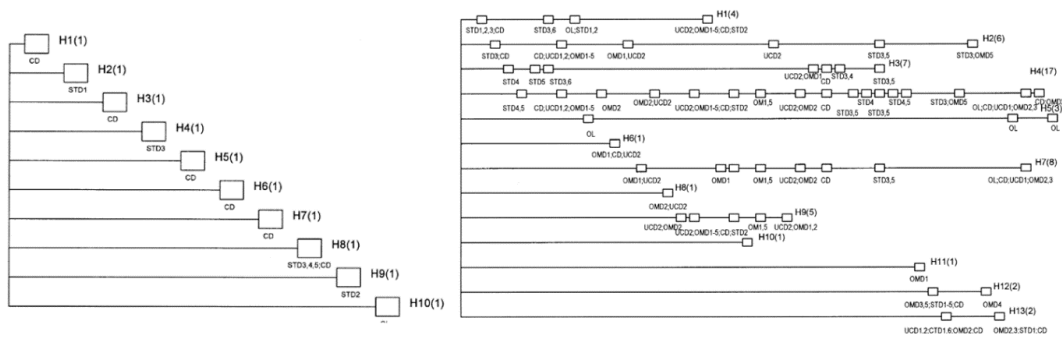


Figure 24 Hypothesis Behavior Graph of Control Group (Left) and Experimental Group (Right) (Kim, Hahn, and Hahn 2000)

In another study (Masri, Parker, and Gemino 2008), the cognitive integration process is included in the process comparison. There are three major categories for the comparative research of conceptual modelling techniques: product comparison (modelling effectiveness), process comparison (modelling efficiency), and understanding-level comparisons (readability efficiency). This study focuses on the third category, investigating the effectiveness of modelling techniques from a problem-solving (understanding) perspective. Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML) are used for theoretical support.

Cognitive Load Theory (CLT) (Sweller 1988) defines the cognitive load with two sources: intrinsic and extraneous. Intrinsic cognitive load is directly related to the interactivity of elements within the task being learned, involving schema acquisition and knowledge construction by combining new information with prior knowledge. Massive elements that must be assimilated simultaneously in one task can increase the intrinsic load on working memory. Extraneous cognitive load involves the process of manipulation of the elements of the message to construct knowledge. For instance, locating or arranging elements from a conceptual model belongs to extraneous cognitive load. The element definition here can vary depending on the subject's previous domain knowledge. Subjects familiar with the tasks can treat several elements as a whole, while those unfamiliar with the tasks need to process elements as many as they are, which can increase the workload.

Another theory used in the study (Masri, Parker, and Gemino 2008) is the Cognitive Theory of Multimedia Learning (CTML). Mayer (Mayer 2001) developed CTML based on findings from various empirical research studies. CTML is based on three assumptions:

- Dual channels: According to dual coding theory (Paivio 1990), individuals have two separate channels to interpret visual and auditory information.
- Limited capacity: According to the working memory theory (Baddeley 1992), individuals cannot treat information from each working memory channel.

- Active processing: According to generative theory (Wittrock 1989), individuals are active processors instead of passive processors. Dynamic processors mean that the individuals “pay attention, organise incoming information, and integrate the information with knowledge stored in long-term memory (prior knowledge)”.

CTML aims to use multimedia presentations to reduce overall cognitive load by reducing extraneous cognitive load. Mayer defines multimedia as “*the presentation of material using words and pictures*”. Authors (Masri, Parker, and Gemino 2008) claim that “*according to the sensory modality description, a textbook with pictures would be considered multimedia as readers will visually process pictures and convert words into sounds for verbal processing (auditory processing)*”. Figure 25 shows an overview of CTML.

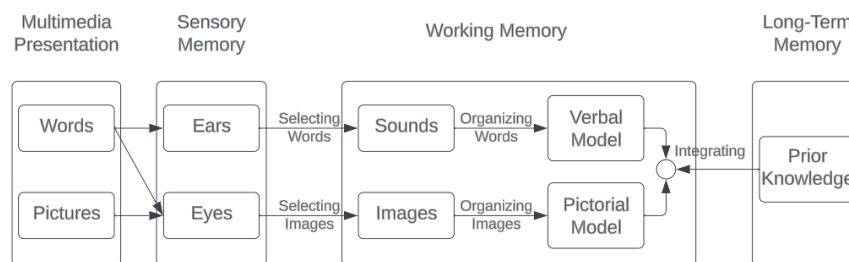


Figure 25 The Cognitive Theory of Multimedia Learning (Mayer 2001)

Based on CTML, Mayer proposed seven design principles to help designers design effective multimedia presentations (see Table 2).

Table 2 Seven principles of CTML (Masri, Parker, and Gemino 2008)

Design Principal	Description
Multimedia Principle	Recipients learn better from words and pictures than from words alone.
Spatial-Contiguity Principle	Recipients learn better when corresponding words and pictures are presented near each other rather than far from each other on a page or screen.
Temporal-Contiguity Principle	Recipients learn better when corresponding words and pictures are presented simultaneously rather than successively.
Coherence Principle	Recipients learn better when extraneous material is excluded rather than included in the presentation.
Modality Principle	Recipients learn better from animation and narration than from animation and on-screen text (spoken text rather than printed text).
Redundancy Principle	Recipients learn better from animation and narration than from animation, narration, and text.
Individual-Differences Principle	Design effects are stronger for low-knowledge learners than high-knowledge learners and for high-spatial learners rather than low-spatial learners.

This study (Masri, Parker, and Gemino 2008) experimented with two cases using the Cognitive Load Theory (CLT) and the Cognitive Theory of Multimedia Learning (CTML) as the theoretical foundation. The results indicate that iconic graphics in the Entity Relationship Diagram significantly impact understanding rather than retention. This study also points out that Previous Domain Knowledge (PDK) and English as a Second Language (ESL) are essential variables to consider in understanding the impact of iconic visual graphics. Previous studies (Khatri et al. 2006; Shaft and Vessey 1998; Mayer 2001; Sweller and Chandler 1994) have indicated the critical difference between experts and novices, with or without consideration of CLT or CTMT. A high level of PDK lowers the intrinsic cognitive load, and design effects will have a lower impact on those with high PDK. The results of this paper about the PDK needed to provide robust evidence of the impact of PDK on the understanding of iconic graphics in ERD. The author speculates that the cause may relate to the instrument used to measure PDK. On the other hand, they include ESL as another covariate to isolate possible effects associated with language processing. Results show a more effective impact in improving the understanding of iconic visual graphics with the non-ESL group. They argue that the results obtained with a non-ESL group are more reliable than those with the ESL group since the tasks in the experiment require language skills, which can bias the results. Working with a foreign language can increase the intrinsic load, which may not be compensated for by the reduction in extraneous cognitive load provided by graphics.

Summary of Diagrammatic Reasoning

This section presents several studies discussing the cognitive process. Each theory has different terms, but from the definition of the concepts, we can notice that additional terms from other theories sometimes describe very similar processes. Figure 26 illustrates the conceptions introduced in this section. To simplify the diagram and the discussion, another diagram with only the visual channel is depicted in Figure 27.

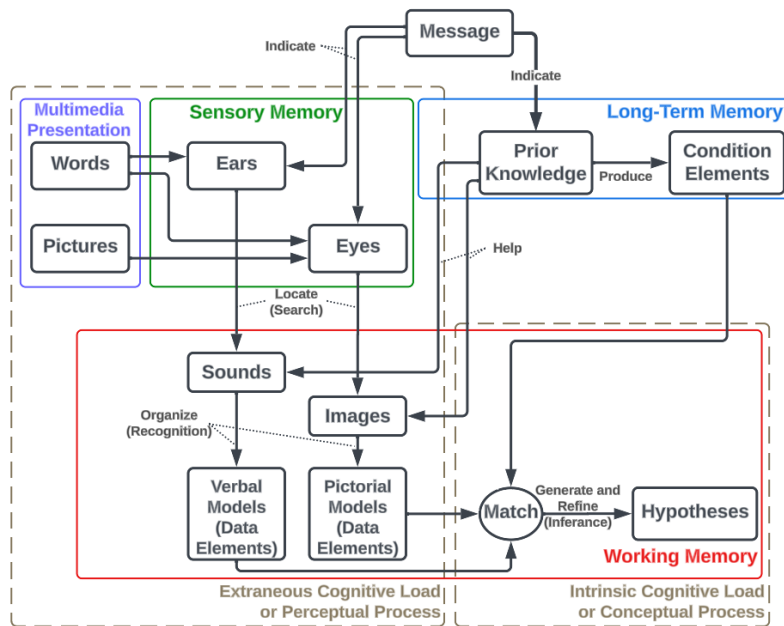


Figure 26 Illustration of the diagrammatic reasoning

The approaches to improving the visual graphics we presented in the section on graphical conceptual modelling notations mainly introduce a positive impact on the perceptual process (or extraneous cognitive load). In other words, those approaches help the subject recognise the perceived visual information more quickly. As mentioned in (Masri, Parker, and Gemino 2008), their study aims to improve the cognitive process by reducing the extraneous cognitive load, thus facilitating other activities within the limited working memory.

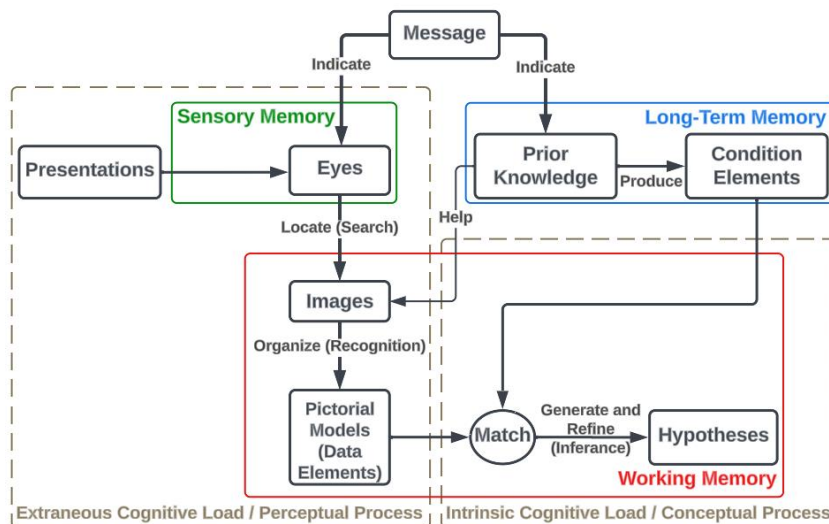


Figure 27 Illustration with the visual system

Furthermore, this also explains why, theoretically, such approaches tend to impact novices significantly more than experts. Experts can rely on diagram schemas stored in their long-term memory, thereby reducing the effort required for interpretation in the working memory, about

either locating elements efficiently or how easily the experts can transform elements into pictorial models to be prepared for knowledge construction. Novices, on the other hand, require more working memory space for the search and recognition process. Therefore, studies proposing new visual graphics are particularly beneficial for novices, as experts already benefit from the support provided by their long-term working memory. However, the difference between novice and expert is not limited to the perceptual process. It also involves a conceptual process. Experts possess problem-solving strategies stored in their long-term memory, helping them generate and refine hypotheses.

The scope of this thesis covers both perceptual and conceptual aspects. However, in the early stage of the research, the impact of our study is primarily on the perceptual side, as we propose the new 3D stereoscopic iconic symbols. As the research continues, we will consider further aspects of the conceptual process, such as placing the information efficiently in the environment or easing the navigation to maximise interface consistency.

2.1.4 Quality of Graphical Conceptual Modelling Notations

Defining quality criteria that objectively assess the effectiveness and utility of graphical conceptual modelling notations is crucial. There are various evaluation approaches such as The Semiotic Quality (SEQUAL) (Krogstie, Sindre, and Jørgensen 2006), Guidelines of Modelling (GoM) (Schuette and Rotthowe 1998), Quality in Conceptual Modelling (Lindland, Sindre, and Sølvsberg 1994), or Seven Process Modelling Guidelines (7PMG) (Mending, Reijers, and Van Der Aalst 2010), etc. In this section, three sets of principles for the design of modelling languages are presented in detail. Those principles apply to the design phase of new graphical conceptual modelling notations or improving an existing graphical abstract modelling notation.

Principles for Modelling Language Design

In a study (Paige, Ostroff, and Brooke 2000), authors proposed nine principles for modelling language design, drawing on object-oriented and mathematical modelling languages. These design principles improve the quality of modelling languages in the software system context, which means that the software created by the modelling language following the principles should have a higher quality. A modelling language is “*used to specify, visualise construct, and document a software system*” (Rumbaugh, Jacobson, and Booch 1999). The designers use it as a tool to support critical and complicated tasks. There are four essential tasks for software designers (Paige, Ostroff, and Brooke 2000):

1. Architectural description: Designers need to describe the abstractions (classes, processes, use cases, programs, etc.) and their relationships

(inheritance/subtyping, data flow, sequencing, etc.) of a system from various levels of detail, which together form the architectural description.

2. Behavioral description: Designers need to specify what each abstraction in the model represents, what each does, and when interactions between model components occur.
3. System documentation: Designers must explain how a system operates and the necessary steps for adapting, maintaining, meeting changing requirements, or correcting mistakes or omissions.
4. Forward and backwards generation: Designers should be able to map architectural abstractions to the program and produce models from the program.

Based on this observation, they proposed the following principles, where they consider simplicity as the main/preemptive principle, along with other fundamental principles that also should be considered:

- Simplicity: No unnecessary complexity is included in the language.
- Uniqueness: There are no redundant or overlapping features.
- Consistency: Language features cooperate to meet language design goals.
- Seamlessness: The same abstractions can be used throughout development.
- Reversibility: Implementation changes can be propagated into the model.
- Scalability: Large and small systems can be modelled.
- Supportability: The language is usable by humans and supportable by tools.
- Reliability: The language encourages the production of reliable software.
- Space economy: Concise models are produced.

The Cognitive Dimensions of Notations

The Cognitive Dimensions of Notations (CDs) framework (Blackwell et al. 2001) has been developed to aid notational system and information artefact designers in evaluating their designs considering the impact on the user-end. The framework highlights the available design choices for such designers, including the characterisation of the user's activity and the inevitable tradeoffs between potential design options. According to the authors, expert designers should produce well-designed information artefacts for users' activity. However, most designers are experts in computer science or engineers. They understand the technical problems they are addressing but are unfamiliar with the problem of understanding at the user end. The authors claim that one solution is to provide a vocabulary that addresses potential design problems, which means "*a vocabulary informed by research in cognitive psychology but oriented toward the understanding of a system developer.*" Unlike other techniques of analysing the usability of computer systems focusing on the finest details of interaction such

as key-press times, visual recognition or memory retrieval, CDs, as a discussion tool for designers, try to provide some perspectives of the interaction with information artefacts at a broad-brush level. The dimensions are presented here with a brief explanation (Blackwell et al. 2001):

- Viscosity: Resistance to change.
- Visibility: Ability to view components easily.
- Premature Commitment: Constraints on the order of doing things.
- Hidden Dependencies: Important links between entities are not visible.
- Role-Expressiveness: The purpose of an entity is readily inferred.
- Error-Proneness: The notation invites mistakes, and the system gives little protection.
- Abstraction: Types and availability of abstraction mechanisms.
- Secondary Notation: Extra information in means other than formal syntax.
- Closeness of Mapping: Closeness of representation to domain.
- Consistency: Similar semantics are expressed in similar syntactic forms.
- Diffuseness: Verbosity of language.
- Hard Mental Operations: High demand for cognitive resources.
- Provisionality: Degree of commitment to actions or marks.
- Progressive Evaluation: Work-to-date can be checked at any time.

The Physics of Notation

The issue of insufficient attention to the graphic visual of conceptual notation has become increasingly prominent in recent years, and this problem persists within the system engineering domain as well. Moody (Moody 2009) proposed a set of principles to design cognitively effective visual notations, i.e., optimised for human communication and problem-solving. The set of principles is named the Physics of Notations (PoN), focusing on the physical (perceptual) properties of notations rather than the logical (semantic) properties. Figure 28 illustrates the scope of the Physics of Notations. The semantic issues about choosing appropriate semantic constructs or defining their meaning are not included in this paper; neither are the sentence-level issues about effectively representing diagrams. The principles are named in the positive sense, representing preferable properties of notations. They can guide the notation's design or evaluate the notation.

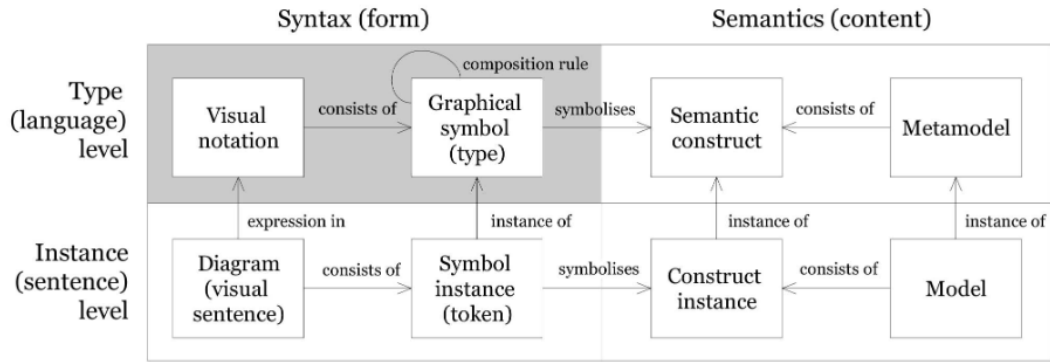


Figure 28 Scope of PoN on the top left-hand quadrant of the diagram (Moody 2009)

There are nine principles with 25 detailed criteria in PoN:

Semiotic clarity: the satisfaction or not of a notation system in Goodman's theory of symbols (Goodman 1976) states that there must be a one-to-one mapping from semantic constructs to visual syntax. One or more of the four anomalies (see Figure 29) can happen when this requirement is violated:

- *Symbol Redundancy:* More than one graphical symbol represents the same semantic construct.
- *Symbol Overload:* One graphical symbol represents more than one semantic construct.
- *Symbol Excess:* Graphical symbols do not correspond to any semantic constructs.
- *Symbol Deficit:* Semantic construct cannot be represented by graphical symbols.

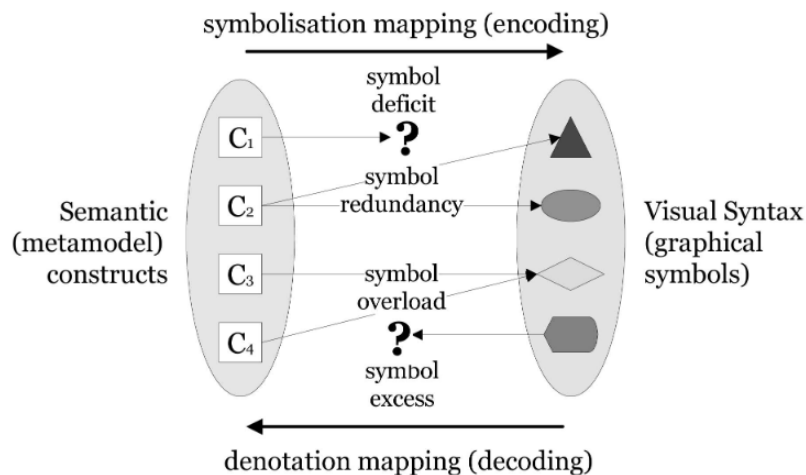


Figure 29 Anomalies of Semiotic Clarity (Moody 2009)

Perceptual discriminability: the ease and accuracy with which symbols can be differentiated.

- *Visual Distance*: Discriminability mainly depends on the visual distance between the symbols. It is measured by the number of visual variables (Bertin 1983) (see Figure 30) that differ in the symbols and the size of the difference.

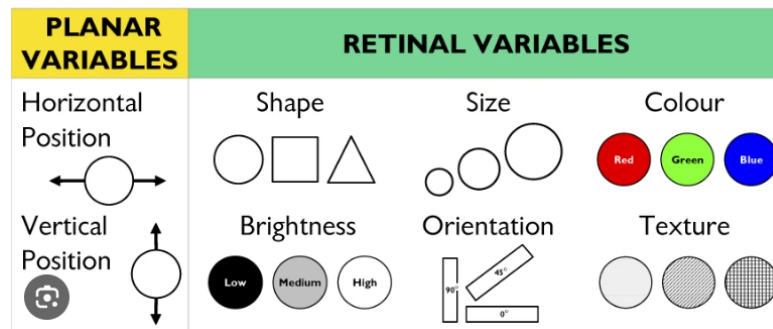


Figure 30 Bertin's eight visual variables (Moody 2009)

- *The Primacy of Shape*: Shape plays a vital role in distinguishing visual variables. Discriminability can be improved using shapes from different families (e.g., round and rectangle variants).
- *Redundant Coding*: Multiple visual variables (e.g., colour and shape) can increase the discriminability.
- *Perceptual Popout*: If the symbol possesses unique values for at least one visual variable, it can be detected attentively and parallel across the visual field.
- *Textual Differentiation*: Text can be used to differentiate the symbols, but it can be cognitively ineffective if it is the only factor that determines the symbols.

Semantic transparency: the extent to which the meaning of a symbol can be inferred from its appearance.

- *Icons (Perceptual Resemblance)*: Icons refer to symbols that are perceptually similar to the semantic constructs, as discussed in the previous section with the semiotic terminology.
- *Semantically Transparent Relationships*: This involves the approaches to represent the relationships. The spatial arrangements of the visual elements can lead to a particular interpretation of the relationships. (see Figure 31)

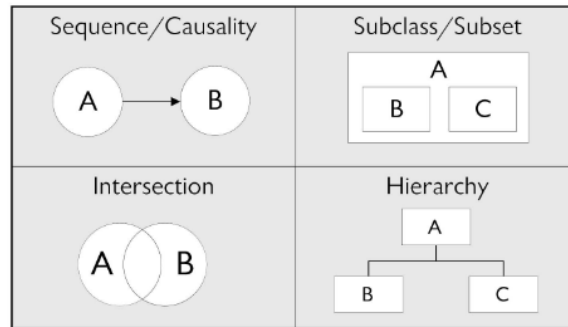


Figure 31 Semantically transparent relationships (Moody 2009)

Complexity management: the ability to represent information without overloading the human mind.

- *Modularization:* It refers to dividing the large systems into smaller subsystems.
- *Hierarchy:* Hierarchy allows a representation of the system with different levels of detail. (see Figure 32)

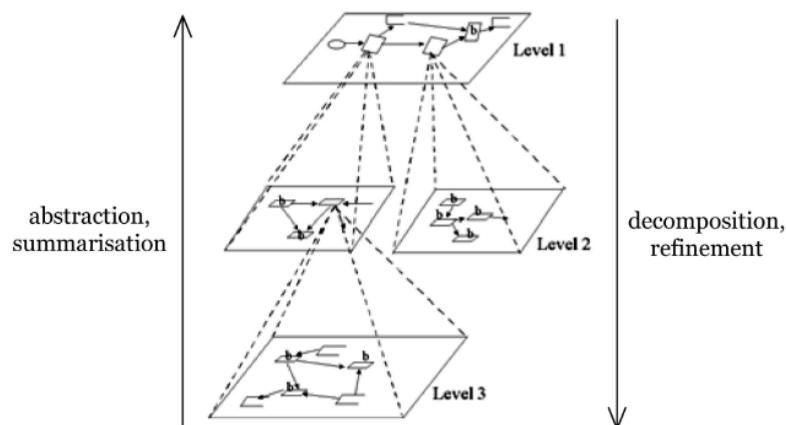


Figure 32 Hierarchical organisation (Moody 2009)

Cognitive integration: the additional cognitive demands on the reader to mentally integrate information from multiple diagrams to represent a system (referencing the study (Kim, Hahn, and Hahn 2000)).

- *Conceptual Integration:* “Mechanisms to help the reader assemble information from separate diagrams into a coherent mental representation of the system.” (Moody 2009)
- *Perceptual Integration:* “Perceptual cues to simplify navigation and transitions between diagrams.” (Moody 2009)

Visual expressiveness: the number of observable variables used in a notation.

- *Use of Color:* Color is considered one of the most cognitively effective visual variables.

- *Choice of Visual Variables:* The choice of the visual variables should be following the content instead of being arbitrary. Table 3 shows the properties of each visual variable where power is the highest level of measurement that can be encoded, and capacity is the number of perceptible steps.

Table 3 Properties of visual variables (Moody 2009)

Variable	Power	Capacity
Horizontal position (x)	Interval	10-15
Vertical position (y)	Interval	10-15
Size	Interval	20
Brightness	Ordinal	6-7
Colour	Nominal	7-10
Texture	Nominal	2-5
Shape	Nominal	Unlimited
Orientation	Nominal	4

- *Textual versus Graphical Encoding:* Graphical encoding is preferred to textual encoding for maximising visual expressiveness. (see Figure 33)

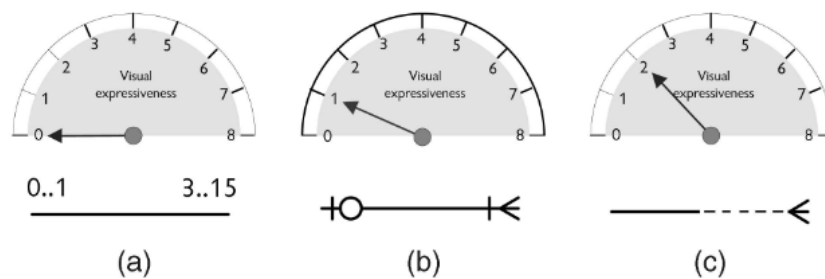


Figure 33 Levels of visual expressiveness: (a) UML=0, (b) IE=1 (shape), and (c) Oracle=2 (shape, brightness) (Moody 2009)

Dual coding: using textual encoding to supplement rather than substitute for graphics since, according to dual coding theory, using text and graphics together to convey information is more effective than using either alone.

- Annotations: Textual explanations for symbols.
- Hybrid Symbols: Graphics and text can be used at the same time to encode the graphical symbols. (see Figure 34)

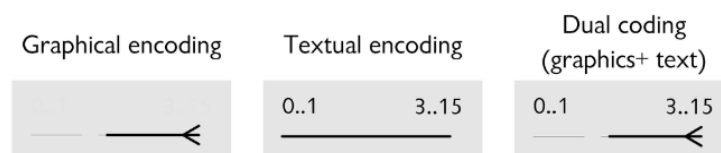


Figure 34 Illustration of dual coding (Moody 2009)

Graphic economy: Limit the number of graphical symbols in a notation to consciously maintain the meanings of symbols in working memory.

- Reduce Semantic Complexity: Graphic economy can be improved by reducing the semantic constructs in one notation.
- Introduce Symbol Deficit: Graphic economy can be improved by reducing the graphical symbols (without reducing the semantic constructs, i.e., causing symbol deficit).
- Increase Visual Expressiveness: Instead of reducing the number of symbols, the graphic economy can be improved by increasing human discrimination ability (i.e., visual expressiveness).

Cognitive fit: Different representations of information are suitable for different tasks and audiences. Visual dialects should be created for such reasons.

- Expert-Novice Differences: Experts and novices process the diagrams in different ways. The diagram schemas in long-term memory developed by the notation experts will largely automate the diagram interpretation process.
- Representational Medium: The visual dialects can vary on the different representational media, such as sketching on whiteboards or paper, rather than computer-based drawing tools.

After reviewing the principles of PoN, it is noticeable that they interact. These interactions can be leveraged to make trade-offs, address conflicts between principles, or exploit synergies where principles support each other. Figure 35 shows an overview of the interactions among the principles.

	Semiotic Clarity	Perceptual Discriminability	Semantic Transparency	Complexity Management	Cognitive Integration	Visual Expressiveness	Dual Coding	Graphic Economy	Cognitive Fit
Semiotic Clarity								±	
Perceptual Discriminability					+			+	
Semantic Transparency		+						±	
Complexity Management							-	+	
Cognitive Integration				+			-		
Visual Expressiveness		+					+	±	
Dual Coding								+	
Graphic Economy		+	+		-			+	
Cognitive Fit									

+ positive effect
± positive effect or negative effect depending on the situation
- negative effect

Figure 35 Interactions between principles (Moody 2009)

Summary of Diagrammatic Reasoning

Three sets of principles are presented in this section, established from a different angle (help designers understand the cognitive process at the user end versus mainly focusing on the physical properties) or in a different context (software system versus systems engineering). As exploratory research of this thesis, one of the main contributions is the new visual metaphors of semantic constructs used in the system engineering activities in an immersive environment. Thus, the PoN is selected as the primary approach to evaluate the visual graphics of our research. However, comparisons between the other two principles and PoN are carried out. The following two tables show the match between principles in Principles for modelling language design and CDs compared to principles in PoN. After the PoN principles, the asterisk indicates that it is not an exact match but shares a certain degree of similarity.

Table 4 Comparison between Principles for modelling language design and PoN

Principles for modelling language design	PoN
Simplicity	Complexity Management + Graphic Economy + Perceptual Discriminability + Semantic Transparency
Uniqueness	Semiotic Clarity*
Consistency	Cognitive Fit*
Seamlessness	Semiotic Clarity
Reversibility	No match
Scalability	Complexity Management
Supportability	No match
Reliability	No match
Space economy	Graphic Economy

Table 5 Comparisons between CDs and PoN

CDs Framework	PoN
Viscosity	No match
Visibility	Complexity Management + Cognitive Integration
Premature Commitment	No match
Hidden Dependencies	Complexity Management + Cognitive Integration
Role-expressiveness	Semantic Transparency + Cognitive Integration
Error-Proneness	Semiotic Clarity + Perceptual Discriminability + Semantic Transparency + Cognitive Integration + Complexity Management
Abstraction	Complexity Management
Secondary Notation	Cognitive Fit*
Closeness of Mapping	Semantic Transparency
Consistency	No match
Diffuseness	Cognitive Integration*
Hard Mental Operations	Cognitive Integration + Complexity Management + Graphic Economy + Perceptual Discriminability + Semantic Transparency
Provisionality	No match
Progressive Evaluation	No match

Moody pointed out that when he conducted this research (Moody 2009), the CDs framework was the closest study as a theory about the visual notation design. He claimed that the goal of

this paper (Moody 2009) is to continue the research and analysis of visual notation, and he referred to CDs as a crucial pioneering work in visual notation. Nevertheless, from the comparison, we can notice that some principles in the CD framework need to be included in PoN, as PoN focuses only on the perceptual part. In contrast, CDs try to provide insights from a broader perspective. For instance,

The same remark can be observed by comparing principles for modelling language design and PoN. Principles for modelling language design are proposed within the software system, while PoN is within the system engineering. However, some principles are still applicable or can be adapted considering the scope of our research. For instance, the adapted version of reversibility can be one principle for our study where the reversibility is no longer between the code and models but between the models in 3D immersive environments and the models in the software for high-level detailed development. This “reversibility” can help the iterative design process between different interfaces. A more thorough investigation into the valuable principles (or their adapted version) within the scope of this thesis can be resigned to future studies.

2.1.5 Validation Method for Graphical Conceptual Modelling Notations

Moody’s Physics of Notation greatly influenced the community as a prominent approach to propose a framework for designing visual notation and principles for the evaluation. However, it does not provide standardised guidelines for the application of principles on the visual notations (Ziehmann and Lantow 2021; El-Attar 2019; Storrie 2013; Van Der Linden, Zamansky, and Hadar 2016), leading to much evaluation with ambiguous results, since each researcher has their interpretation of the application of PoN:

- *“We discover that the PoN, in its current form, is neither precise nor comprehensive enough to be applied objectively to analyse practical visual software engineering notations.”* (Storrie 2013)
- *“The PoN has been used to analyse existing standard visual notations (such as BPMN, UML, etc.) and is commonly used for evaluating newly introduced visual notations and their extensions. However, due to the rather vague and abstract formulation of the PoN’s principles, they have received different interpretations in their operationalisation.”* (Van Der Linden, Zamansky, and Hadar 2016)
- *“PoN theory does not propose recommendations on how the principles should be assessed in terms of notation validation, nor does it define the threshold values above which the principles are met.”* (Kuhar and Polančič 2021)

A study (Van Der Linden and Hadar 2019) conducted a systematic literature review of the PoN’s applications. The study selected 775 papers and ended up with 70 papers. PoN

applications are present among the papers for new, proposed, and existing visual notations. (Van Der Linden and Hadar 2019) The study brought up four research questions, and here is the feedback on those research questions.

- RQ1: Which visual notations have been analysed using the PoN theory?

Among the papers citing the PoN, only 16% are concerned about its application of PoN. Most papers citing PoN mention it as a general reference to indicate that the visual notation or variable is essential. Almost half of the papers with the application of PoN are for entirely new visual notations, and the other half are for previously existing notations.

- RQ2: What reasons do the researchers provide for using the PoN theory?

They found that the alternative approach of PoN is typically the CD approach, followed by SEQUAL. However, 83% of the papers used PoN without justifying anything. Only 14% of papers justify their use of the PoN theory with some explicit arguments. It cited an example where the authors state the reason for the selection of PoN as “for the visual language proposed here, the PoN principles are applied because of their scientific and theoretical validity” (Herter, Brown, and Ovtcharova 2013) or another more general example like “we chose PoN theory as it is the state of the art SE and RE notation evaluation frameworks widely used with other notations.” (Popescu and Wegmann 2014) or as in another paper where the author referred what the Moody discussed about the limits and disadvantages of CDs (Moody 2009) and use it as a justification of choosing PoN. This observation has largely demonstrated the enormous impact of PoN in relevant fields.

- RQ3: To what degree do the analyses consider the requirements of the notation’s users?

This is an interesting and important question since the ultimate goal behind the applications is always to improve the user experience. Surprisingly, the result is the exact opposite. Only 6% possess elicitation of requirements from users. This is noteworthy since PoN is widely criticised for being vague and ambiguous, and involving users might alleviate some ambiguity.

- RQ4: How verifiable are the performed analyses?

Among the nine principles, some are evaluated frequently in papers (some even close to 100%), and conversely, some are barely mentioned. This can reflect the need for more operational support of some principles. For instance, for the principle of semiotic clarity, the notations are measured by Goodman’s theory of symbols. (Goodman 1976) There should be a 1:1 correspondence between semantic constructs and graphical symbols. When it is not respected, one or more anomalies can occur: symbol redundancy, symbol overload, symbol excess, or symbol deficit. With the support of theory and explicit guidance, uncertainty and

ambiguity are eliminated. Hence, the coverage of usage of this principle reaches 100% among the papers analysing existing notations. However, this is not always the case for the other principles. The principles such as cognitive fit, cognitive integration, and complexity management seem to be reported much less than others. For the same reason, when each principle is mentioned, those with more solid support are discussed more explicitly than others without it. Furthermore, the gap exists among principles and depends on the language applied. The applications of PoN to new notations show a markedly smaller scope in terms of detailed and verifiable reporting. This indicates that applying the PoN to guide the design of new notations is even farther from the ideal scenario than applications that evaluate existing notations.

After gaining insight into the background of PoN application with the study (Van Der Linden and Hadar 2019), several studies involving the application of PoN will now be presented.

(Anwer and El-Attar 2014) used PoN to evaluate the statechart diagrams, one of the Unified Modelling Language (UML) pillar diagram types. Due to the paper limit, they presented the evaluation of statechart diagrams with Semiotic Clarity, Visual Expressiveness, Perceptual Discriminability, Dual Coding, Complexity Management, and Graphic Economy principles. Their main approach to applying PoN is their observation of the state chart based on their interpretation of the principles.

[Semiotic Clarity] They identified Symbol Redundancy and Symbol Overload by evaluating the statechart by searching through the statechart diagrams' semantic constructs and symbols to verify the one-to-one mapping rule (see Figure 36).

Semantic construct	Symbol
State	
Composite state	
Junction point	

Symbol	Semantic Construct
	<ul style="list-style-type: none"> • State • Composite state • Process state
	<ul style="list-style-type: none"> • Composite state • State
	<ul style="list-style-type: none"> • Junction point between states • Start symbol
	<ul style="list-style-type: none"> • History state • Deep history state • Entry point

Figure 36 Symbol redundancies (left) and Symbol overload (right) (Anwer and El-Attar 2014)

[Visual Expressiveness] Based on Bertin's definition of visual variables (Bertin 1983), they claim that statechart diagrams use two visual variables: shape with four values: rectangle with

rounded corners, circle and square-to-kite, and brightness with two values: white and black. According to this observation, they conclude that statechart diagrams are not visually expressive with the low utilisation of the visual vocabulary.

[Perceptual Discriminability] Statechart diagrams use only two visual variables: shape and brightness. They evaluate the Perceptual Discriminability primarily from the shape, as Moody claimed that shape plays an important role in the visual variables to distinguish. The approach is to compare the symbols belonging to the same shape category, and the result is presented in Figure 37. Based on this result, they state that circle shape symbols do not have a significant visual distance between them, similar to the rectangle shape symbols, resulting in low discriminability.










Circle Shape	
 History	 Dynamic junction point
 Sync state	 End point
 Entry point	 Junction point
 Start symbol	 End symbol
Rectangle Shapes	
 <p>States and Composite States</p>	

Figure 37 Comparison of a circle shape and Rectangle shape (Anwer and El-Attar 2014)

[Dual Coding] They conclude: “The statechart diagrams notation does not utilise dual coding.”

[Complexity Management] They make a statement about Complexity Management without giving an explicit conclusion: “Statechart diagrams notation has explicit mechanisms for complexity management using superstates. Superstates are special states that can contain other statechart diagrams or states”.

[Graphic Economy] They claim that statechart diagram notation contains a more extensive notational set, by Moody’s statement that a visual language should not have more than six symbols.

The same author later published another study (El-Attar 2019) evaluating the visual syntax of use case diagrams. In this study, he provides a more explicit explanation of the operationalisation of each principle in PoN. However, it still needs to be more specific, with an explicit conclusion about the evaluation with PoN principles, and it largely depends on subjective interpretation.

(Storrie 2013) propose an operationalisation approach of Semiotic Clarity and Perceptual Discriminability. They try to propose metrics with adjustable weights to accommodate future empirical findings. The metrics are normalised to range from 0 to 1 and stable, i.e., “the small changes in a notation will lead to relatively small changes in the assessment”. For instance, Figure 38 shows the metrics of the anomalies of Semiotic Clarity: Symbol Redundancy (SR), Symbol Overload (SO), Symbol Excess (SE), and Symbol Deficit. It allows an automated calculation with the input of the sets of concepts and graphemes and their relationships. However, the result can still differ depending on restrictions and assumptions about the notation to be analysed. For example, assuming that several graphical symbols can present the semantic construct actor will lead to symbol redundancy. Still, assuming that the actors shown by different graphical symbols are considered as other semantic constructs will not lead to symbol redundancy.

$$\begin{aligned}
 SR(N) &:= \frac{1}{|visualized(C_N)|} \sum_{c \in C_N} redundancy(c) \\
 SO(N) &:= \frac{1}{|meaningful(G_N)|} \sum_{g \in G_N} overload(g) \\
 SE(N) &:= 1 - \frac{|meaningful(G_N)|}{|G_N|} \quad SD(N) := 1 - \frac{|visualized(C_N)|}{|C_N|}
 \end{aligned}$$

Figure 38 Equations to calculate the metrics of Semiotic Clarity

Another approach to evaluate the principles is from the perceived aspect of the user end. (Figl and Derntl 2011) Uses the principles of PoN as perceivable cognitive effectiveness criteria to explore how they influence perceived usefulness. In their experiment, they showed a set of diagrams to the participants (see Figure 39) and gave them a questionnaire to fill out. Participants are presented with statements regarding the diagrams based on the principles and are asked to rate their agreement with each statement about the diagrams they are evaluating. For instance, there are two statements about Perceptual Discriminability:

- some symbols are difficult to differentiate.
- some symbols can easily be confused with each other.

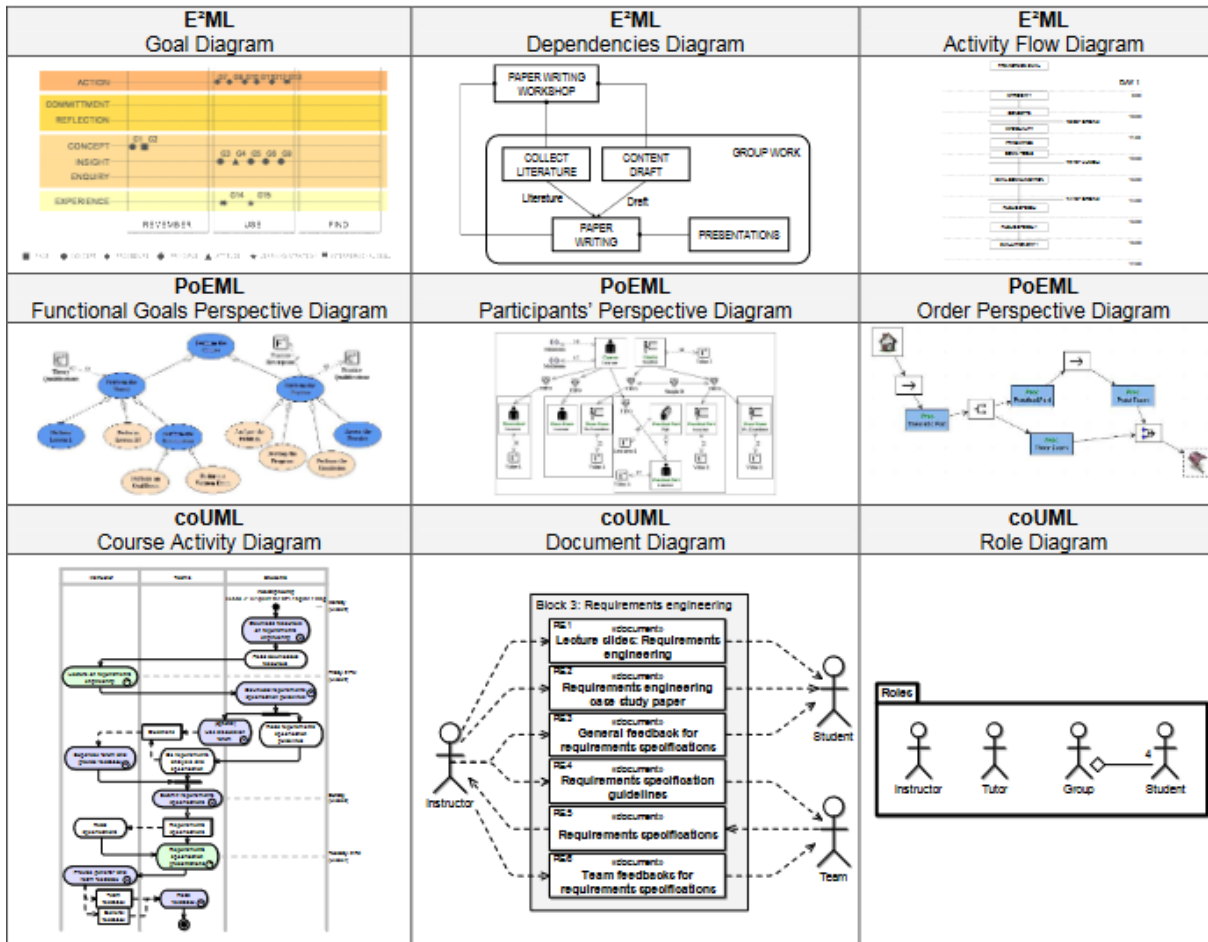


Figure 39 Diagrams used in the user evaluation (Figl and Derntl 2011)

Summary of validation method

The PoN is widely used to evaluate visual notation. However, each study interprets the evaluation approach differently due to insufficient operationalisation support. Moreover, there are no threshold values above which the principles are met. This makes the evaluation impractical with subjective results.

2.1.6 Discussion and Limits

Generally, conceptual modelling notations are mainly for software engineering rather than systems engineering. Very few studies exist on the modelling notation of MBSE (e.g., SysML). There is no clear link between the graphical modelling notation quality criteria and the cognitive processes (and therefore the profile/level of experience of the target users) presented above, which rationalises the design of these graphical conceptual modelling notations. With these ergonomic recommendations or guidelines, we can go through multiple iterations between the design of a visual notation and user testing until we find the notation design that meets the needs and fits the profiles of the target users. Furthermore, the type of graphical conceptual modelling notations is often limited to 2D symbolic visual graphics. Few studies explore

alternative types, such as 3D or iconic, let alone studies that compare the various types of graphical conceptual modelling notations.

2.2 Model-Based Systems Architecting

A graphical conceptual modelling notation is a tool to support users in their activities. This section will review the main concepts, techniques, and tools related to the activities that belong to our topic of interest: systems architecting, especially using a model-based approach.

2.2.1 Systems Architecting

Terms such as “engineering”, “system”, or “architecting” have been in use for a long time before the concept of combining them. From the Oxford English Dictionary, there are definitions for Architect, Engineer, and System:

- Architect: A skilled professor of the art of building.
- Engineer: A person who makes engines, structures or systems.
- System: An organised or connected group of objects.

The first time the term systems engineering was brought up can be traced back to Bell Telephone Laboratories in the 1940s when there was a need to identify and manipulate the properties of a system as a whole. It implies that we may encounter situations where the properties of a whole system can significantly differ from the properties of the sum of the parts. Those needs motivated the industries, especially the US military when they were in the development stage, to apply the systems engineering approach. The International Council on Systems Engineering (INCOSE) defines systems engineering as “*an interdisciplinary approach governing the total technical and managerial effort required to transform a set of stakeholder needs, expectations and constraints into a solution and to support that solution throughout its life*”. In the 1990s and beyond, there was a shift towards architecture-centric approaches within systems engineering. System architecture emphasises the importance of system architecture, a system's high-level structure and organisation. Rehtin (E. Rehtin 1991) states, “*The architect, therefore, is not a ‘general engineer’, but a specialist in reducing complexity, uncertainty and ambiguity to workable concepts. The systems engineer, in contrast, is the master of making feasible concepts work. Architecting is working for a client and with a builder, helping determine the preferred architecture, that is, helping determine relative requirement priorities, acceptable performance, cost, and schedule — taking into account such factors as technology risk, projected market size, likely competitive moves, economic trends, political, regulatory requirements, project organisation, and the appropriate ‘ilities’ (availability, operability, manufacturability, survivability, etc.). Toward the end of the project, architecting also certifies completion and satisfactory operation of the system*”. Three basic worldviews

exist regarding the relationship between systems engineering and systems architecting (Emes et al. 2012).

- One approach perceives systems architecting as “*simply a rebranding of systems engineering to broaden its appeal with no change in content*”.
- The second approach perceives “*systems engineering restricted to its traditional processes, with systems architecting adding to systems engineering through external processes*”.
- The third approach, the most popular one in the community surveyed, perceives “*systems architecting addressing shortcomings in traditional sequential lifecycle models by stretching the content of systems engineering to include new elements under the banner of system architecting*”.

Moreover, Maier and Rechtin (Maier and Rechtin 2000) proposed the difference between engineering and architecting: “*Generally speaking, engineering deals almost entirely with measurables using analytic tools derived from mathematics and the hard sciences; that is, engineering is a deductive process. Architecting deals largely with unmeasurable using nonquantitative tools and guidelines based on practical lessons learned; that is, architecting is an inductive process.*”

Compared to conceptual design or embodiment design, systems architecting is internationally standardised. According to (ISO/IEC/IEEE 42010 2022), architecting involves conceiving, defining, expressing, documenting, communicating, and certifying the proper implementation, maintenance, and improvement of architecture throughout the life cycle of an entity of interest. Architecture refers to fundamental concepts or properties of an entity in its environment and governing principles for the realisation and evolution of this entity and its related life cycle processes. Four lifecycle processes are presented in (ISO/IEC/IEEE 15288 2023) (see Figure 40), where the scope of this thesis is business or mission analysis process, stakeholder needs and requirements definition process, system requirements definition process, and system architecture definition process.

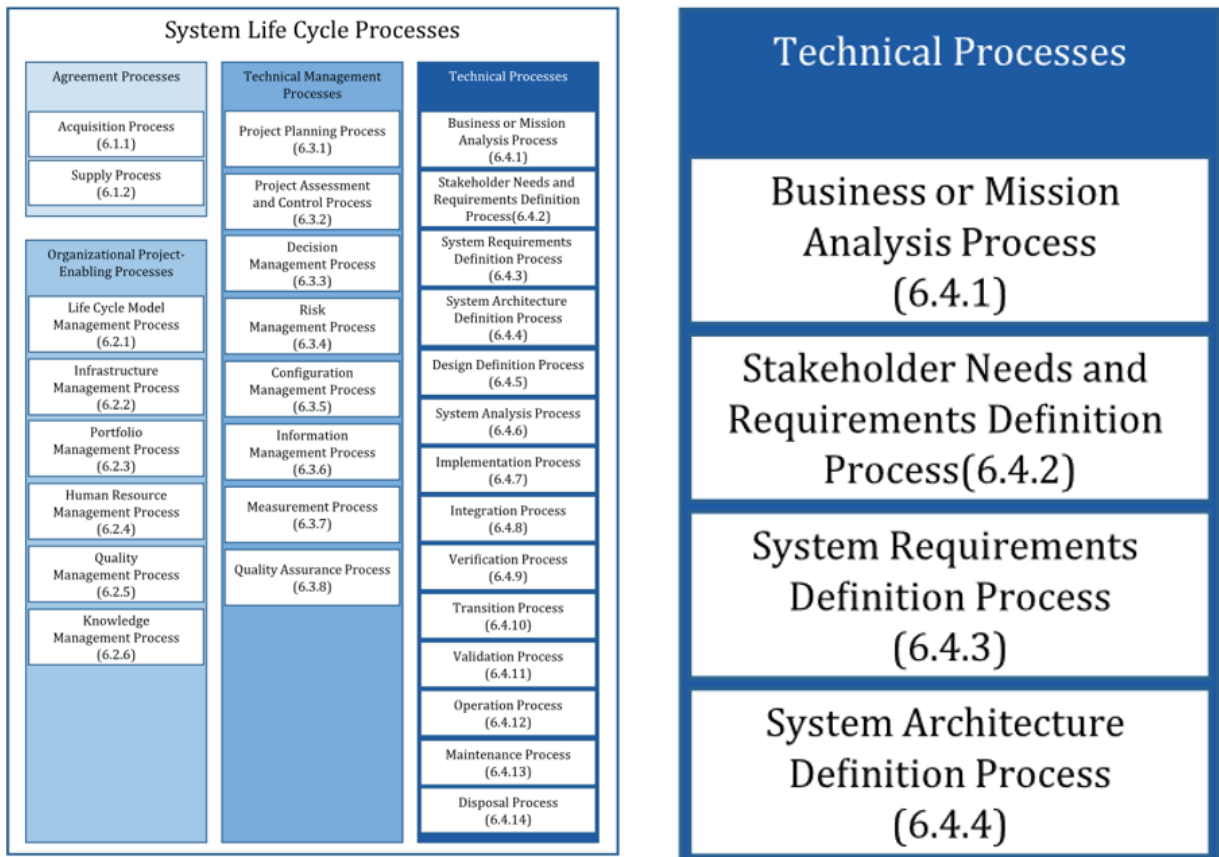


Figure 40 System lifecycle processes (left) and the scope of this thesis (right) (ISO/IEC/IEEE 15288 2023)

Architecture description is an expression of architecture that serves to capture architectural data. According to (ISO/IEC/IEEE 42010 2022), “*architecture descriptions are work products resulting from architecting efforts. As a work product, an AD is devised for the specific purpose of the architecting effort, distinct from the purpose of the entity of interest.*” The definition of architecture descriptions relies on an architecture description framework. An architecture description framework is defined as “*conventions, principles and practices for describing architectures established within a specific application domain or community of stakeholders*”(ISO/IEC/IEEE 42010 2022). Figure 41 illustrates the conceptual model of an architecture description framework.

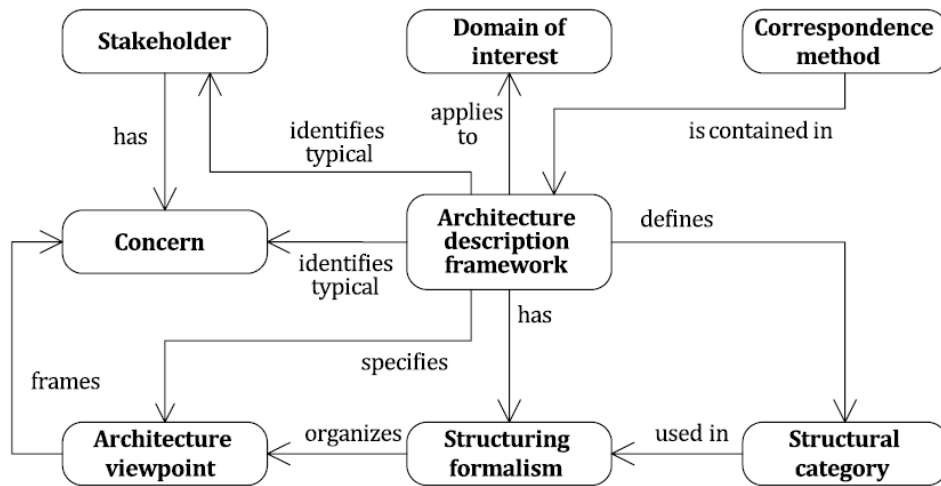


Figure 41 Conceptual model of an architecture description framework (ISO/IEC/IEEE 42010 2022)

Architecture description also requires an architecture description language. An architecture description language is the “*means of expression, with syntax and semantics, consisting of a set of representations, conventions, and associated rules intended to describe an architecture*” (ISO/IEC/IEEE 42010 2022). Figure 42 illustrates the conceptual model of ADL.

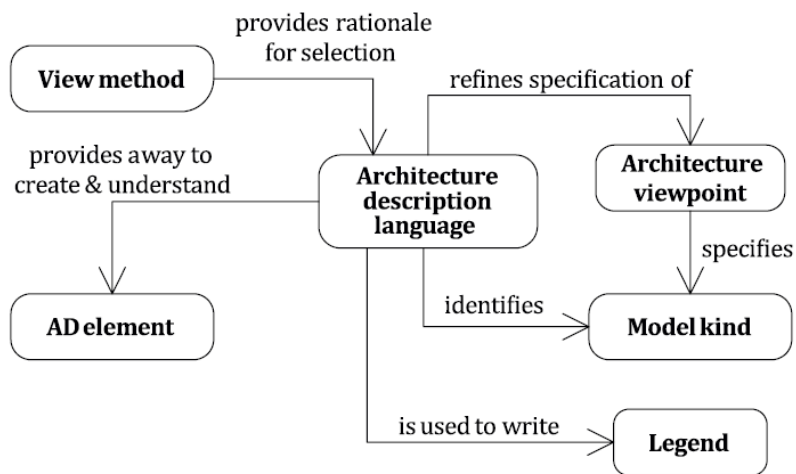


Figure 42 Conceptual model of an architecture description language (ISO/IEC/IEEE 42010 2022)

2.2.2 Model-Based Systems Architecting

As previously introduced, Model-Based Systems Engineering (MBSE) is proposed to address the issues of complex projects with countless text files as a paradigm shift from document-centric to model-centric. It is defined as: “*The formalised application of modelling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle*”

phase” (Friedenthal, Griego, and Sampson 2009). The essential motif is to convert the systems engineering approach from “document-centric” to “model-centric” for the benefits such as better communication, better analysis capability, improved system understanding, improved consistency, reduce time, increased efficiency, better manage complexity, reduce cost, reduce errors, increased traceability, etc. (Campo et al. 2023). Model-Based Systems Architecting (MBSA), often used in place of MBSE, is the formalised application of modelling to support the top-down recursive definition of expected functions, imposed design constraints, requirements and abstract logical architecture, and the continuous early verification, validation, and traceability of architecture descriptions. As Systems Architecting is the front end of Systems Engineering, MBSA is the front end of MBSE.

MBSA aids in designing and developing complex systems by utilising modelling frameworks, languages, and software. (ISO/IEC/IEEE 42020 2019) defines architecture framework as “*conventions, principles and practices for use by architecture-related activities established within a specific domain of application or community of stakeholders.*” This definition expands the definition of architecture framework in (ISO/IEC/IEEE 42010 2022), where it separates the framework as “architecture description framework” and “architecture evaluation framework”. “*Architecture description frameworks (ADF) are used to codify the conventions and common practices of architecture description. Architecture description languages (ADL) are used to codify the description of architectures within different communities and domains of application*”(ISO/IEC/IEEE 42010 2022). ADL is considered a subtype of the conceptual modelling language. The following subsections will introduce some popular cases respectively.

2.2.2.1 MBSA Frameworks

MBSA utilise MBSE frameworks to guide the architectural modelling process. Several mainstream MBSE frameworks exist, especially systems engineering processes and MBSE methods.

ISE&PPOOA

Integrated System Engineering and PPOOA (ISE&PPOOA) (Fernandez and Hernandez 2019) provide an integrated process, methods, and tools for the systems engineering of software-intensive mechatronic systems. It consists of two parts: ISE and PPOOA.

The ISE part includes the first steps of a systems engineering process applicable to any system, not only the software-intensive ones. The ISE sub-process integrates traditional systems engineering best practices with MBSE. There are four steps to elaborate on the ISE sub-process: 1) Identify operational scenarios; 2) Identify system capabilities and HLR /

Specify quality attributes and system NFRs; 3) Create system functional architecture; 4) Create system physical architecture.

The PPOOA (Processes Pipelines in Object Oriented Architectures) part of the process emphasises the modelling of concurrency as early as possible in the software engineering part of the integrated process. The figures below illustrate the sub-process of each part.

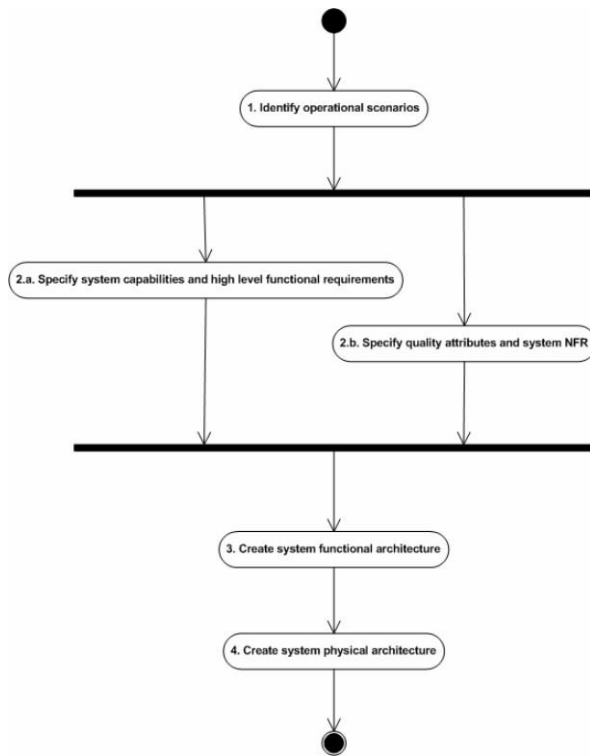


Figure 43 ISE Subprocess (Fernandez and Hernandez 2019)

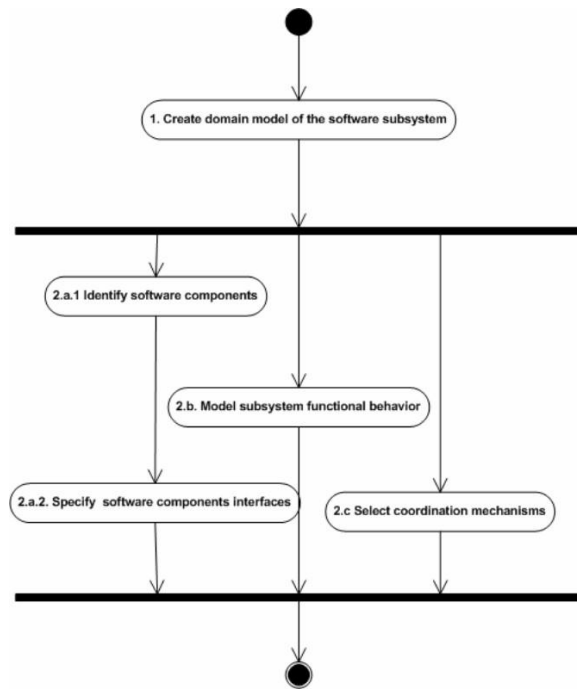


Figure 44 PPOOA Subprocess (Fernandez and Hernandez 2019)

SYSMOD

According to its creator (Weilkiens 2020), SYSMOD is more a discovery than an invention since it consists of well-known methods and practices. He collected practices, transferred some from software engineering to the systems engineering discipline, and described the links between the practices to combine them into a methodology. Customisations of SYSMOD are commonly used in industrial projects, and the author keeps updating the toolbox with modifications based on feedback from user experiences.

The SYSMOD toolbox consists of three main artefact kinds:

- The SYSMOD products are crucial artefacts of the systems development, like requirements or architectural descriptions.
- The SYSMOD Methods are best practices for creating a SYSMOD product.

- The SYSMOD Roles are work descriptions of a person. A role is responsible for SYSMOD Products and is a primary or additional performer of SYSMOD Methods.

SYSMOD is a user-oriented approach to requirements engineering and system architectures. The two figures below demonstrate the processes of analysing requirements and architecture design.

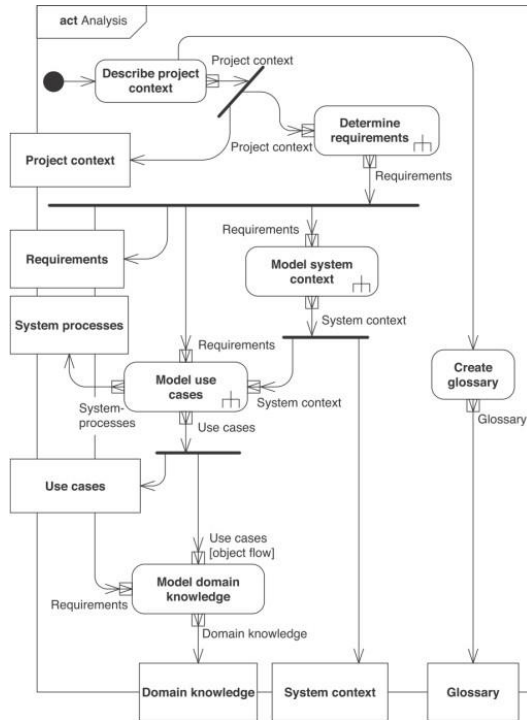


Figure 45 Analysing requirements
(Weilkiens 2020)

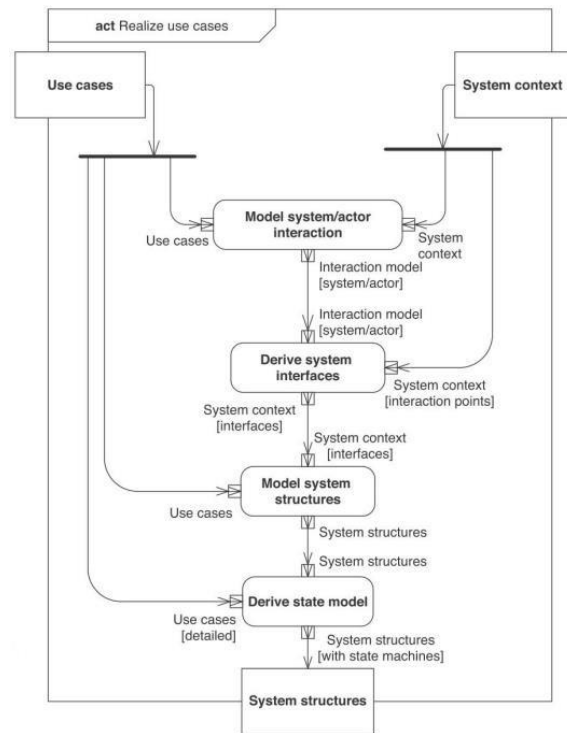


Figure 46 Architecture Design
(Weilkiens 2020)

Harmony-SE

Harmony-SE mirrors the top-left part of the classic V model (Forsberg and Mooz 1991) for system engineering, which includes requirements analysis, system functional analysis and design synthesis (Hoffmann 2014). The key objectives of Harmony for Systems Engineering are 1) Identification and derivation of required system functions; 2) Identification of associated system modes and states; and 3) Allocation of the identified system functions and modes/states to a subsystem structure. Figure 47 depicts an overview of Harmony-SE.

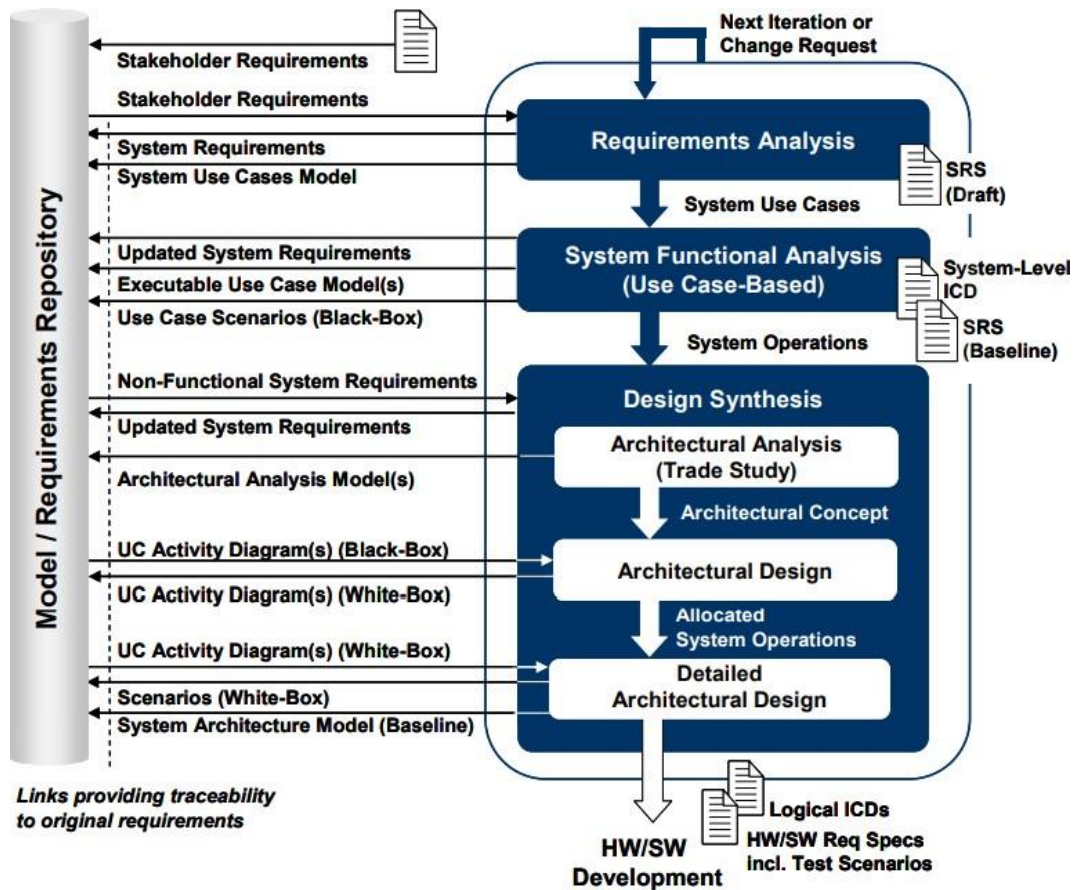


Figure 47 Overview of Harmony-SE (Forsberg and Mooz 1991)

SE-READ

The SE-READ method (Mhenni et al. 2014) is dedicated to mechatronic systems architecture design. Mechatronic systems are multi-disciplinary systems consisting of different subsystems or components from mechanics and electronics interacting together. The design should satisfy system-level requirements or functionalities with the interactions between components since the traditional way of splitting design into separate disciplines is no longer suited for this situation.

This method contains two parts (see Figure 48): the black box analysis as an operational view from the outside that carries a comprehensive and consistent set requirement, and a white box analysis that progressively leads to the internal architecture as a functional view and the behavioural view of the system. The blackbox analysis comprises Global mission, life cycle, context, external interfaces, user operating modes, services provided by the system, functional scenarios, requirements, and traceability. The white box analysis comprises activity diagrams, logical breakdown and allocation, requirements traceability, logical architecture, parametric diagrams, physical allocation, and physical architecture. Each phase has its own proper SysML diagrams to represent it.

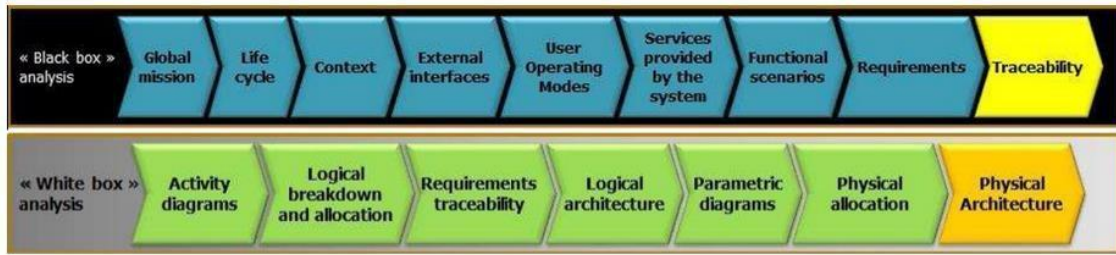


Figure 48 Overview of SE-READ (Mhenni et al. 2014)

OOSEM

The object-oriented systems engineering method (OOSEM) (Friedenthal, Moore, and Steiner 2014; Walden et al. 2015) is a model-based approach to system engineering that integrates top-down and model-based methods. The approach uses OMG SysML to support system specification, analysis, design, and verification. OOSEM leverages object-oriented concepts, traditional top-down systems engineering methods, and other modelling techniques to help build more flexible and scalable systems that adapt to evolving technologies and changing requirements. It also integrates with object-oriented software development, hardware development, and testing.

OOSEM started using UML by non-standard customisation to identify modelling artefacts. Since 2006, OOSEM has adopted the SysML language, and the support of the tool has been significantly improved. The method includes the following activities: 1) Analyze Stakeholder Needs; 2) Define System Requirements; 3) Define Logical Architecture; 4) Synthesize Candidate Allocated Architectures; 5) Optimize and Evaluate Alternatives; and 6) Validate and Verify System.

ASAP

The method ASAP (Guillermo and Góngora 2014) is the adopted MBSE methodology in the railway sector. Railway car builders need a more structured design process to address the new railway operators' lack of technical background. The rolling stock systems required a higher level of complexity to manage the functionalities and services that needed to be implemented, with the constraint of the software. Therefore, Alstom proposed the Advance System Architect Program (Figure 49) as the new Alstom System engineering approach to reduce the “problem” complexity.

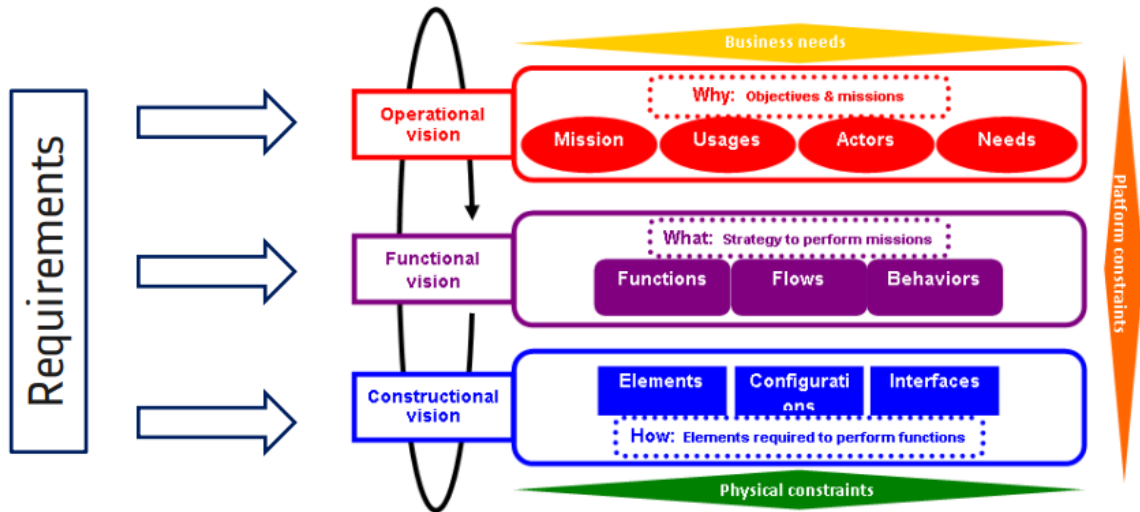


Figure 49 Overview of ASAP (Guillermo and Góngora 2014)

SysCARS

The method SysCARS was proposed by Valeo in 2010 (J.-D. Piques and Andrianarison 2012). Their approach was motivated by the increasing complexity of automotive system design and the need for efficiency and flexibility in a highly competitive context. The process is divided into four major phases: 1) Stakeholder needs definition, 2) Requirements analysis, 3) Logical architecture design, and 4) Physical architecture design.

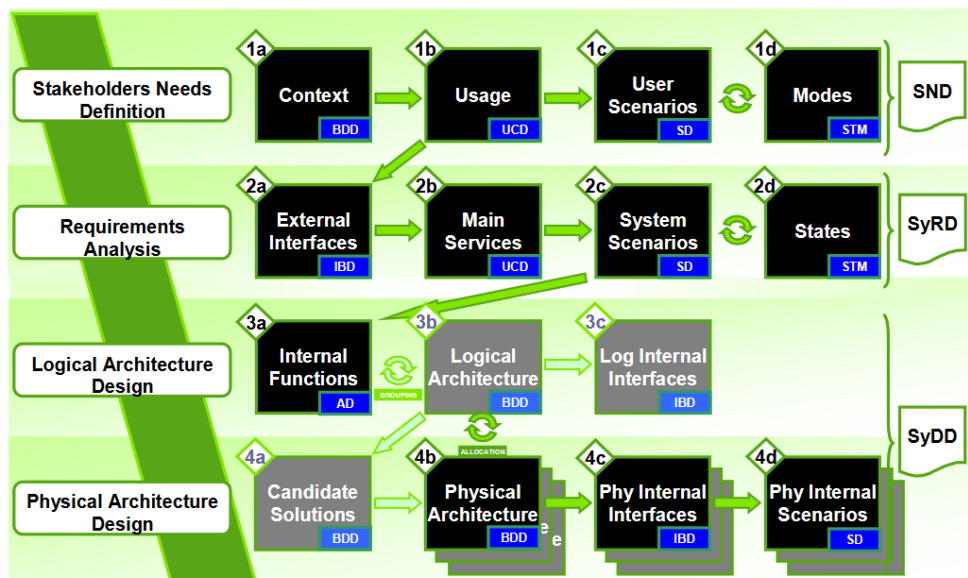


Figure 50 Overview of SysCARS (J. Piques 2014)

ARCADIA

Architecture Analysis & Design Integrated Approach (ARCADIA) (Voirin 2018) is a tool method dedicated to systems and architecture engineering. ARCADIA emphasises the understanding of customer needs, defines and shares the product architecture among all engineering stakeholders, enables early design validation, and facilitates integration, validation, verification, and qualification. ARCADIA has four key phases, illustrated in Figure 51.

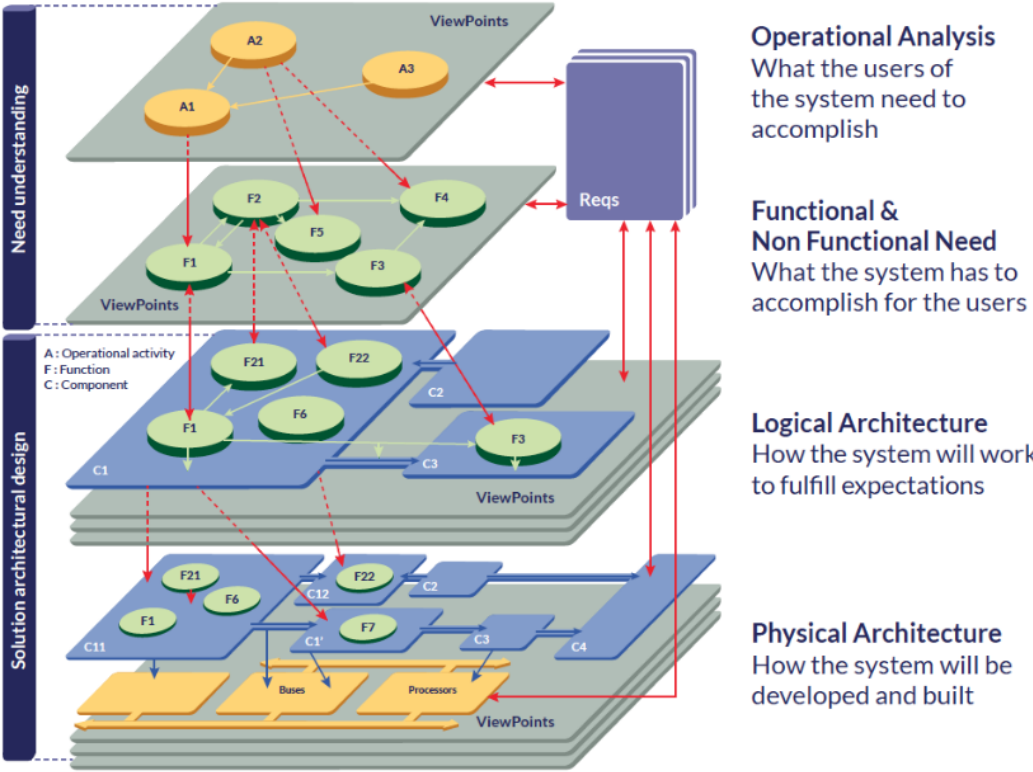


Figure 51 Overview of ARCADIA (Voirin 2018)

2.2.2.2 MBSA Languages

Several popular architecture description languages create the architecture descriptions of an MBSA framework. Many methods adopt SysML—often a tailored version with stereotypes—as the modelling language. This subsection will present SysML and OPM.

SysML

SysML (ISO/IEC 19514 2017) is a modelling language created as a profile of the Unified Modelling Language (UML). It was requested as UML for systems engineering (OMG 2003). It, therefore, reused a large part of UML 2.5 (see Figure 52) with a few changes to attempt to include SE considerations (OMG 2019).

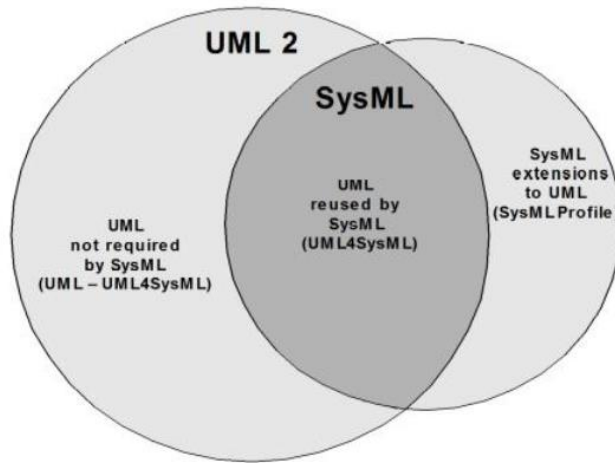


Figure 52 Overview of SysML/UML interrelationship (OMG 2019)

Most SysML diagrams reuse strictly UML diagrams, such as sequence, state machine, use case, and package diagrams. Activity, block definition, and internal block diagrams are modified for consistency with SysML extensions. Parametric and requirement diagrams are new SysML diagram types. A parametric diagram “describes the constraints among the properties associated with blocks” (OMG 2019). A requirement diagram “provides a modelling construct for text-based requirements and the relationship between requirements and other model elements that satisfy or verify them” (OMG 2019). An overview of the SysML diagrams is illustrated in.

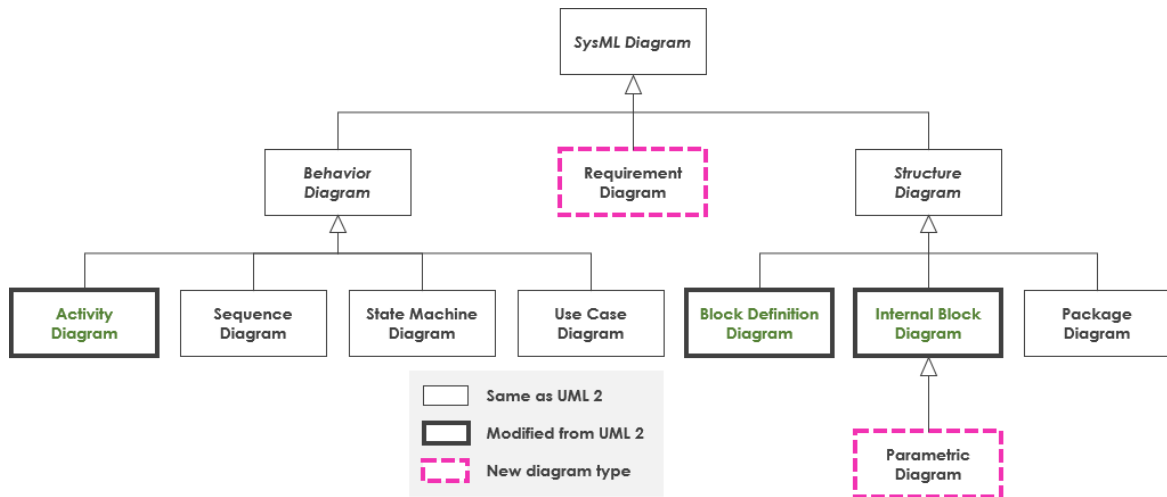


Figure 53 SysML diagram taxonomy (OMG 2019)

OPM

An alternative to the SysML-based method is the Object-Process Methodology (OPM) (Dori 2002). OPM is an MBSE method and MBSE conceptual modelling language for capturing knowledge and designing systems. It combines the object-oriented and process-oriented

paradigms into a unified frame of reference. System structure and behaviour aspects are represented in the same OPM view without prioritising one over the other. The fundamentals of OPM are entities, including Objects that are physical or informatical things, and Processes that are things that transform objects and links:

- “Structural links express static relations between pairs of entities” (Dori, Reinhartz-Berger, and Sturm 2003).
- “Procedural links connect entities to describe the behaviour of a system” (Dori, Reinhartz-Berger, and Sturm 2003).

The behaviour can be presented in three ways (Dori, Reinhartz-Berger, and Sturm 2003): 1) Processes can transform objects, 2) Objects can enable processes, and 3) Objects can trigger events that invoke processes. Furthermore, OPM is bimodal, as graphic and textual are two semantically equivalent modalities (see Figure 54).

- Object-Process Diagram (OPD): OPD is the sole type of diagram utilised in OPM, constituting the graphical, visual OPM formalism.
- Object-Process Language (OPL): OPL is a dual-purpose language; in other words, it is readable by humans and interpretable by computers.

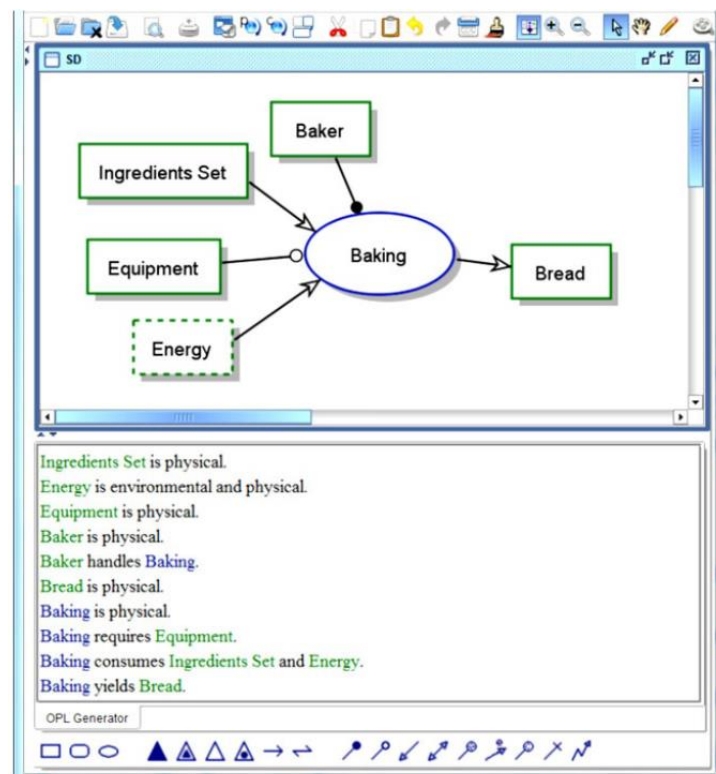


Figure 54 GUI of OPCAT shows the baking system's (SD, top-level) system diagram. Top: object–process diagram (OPD). Bottom: The corresponding, automatically generated Object–Process Language (OPL) paragraph (Dori 2011)

2.2.2.3 MBSA Software

An MBSA software supports the MBSA frameworks by efficiently providing an architecture description language interface. As SysML is an extended version of UML for systems engineering, the software supporting SysML is a software supporting UML and often other conceptual modelling language (e.g., BPMN) that was extended with new features, such as papyrus, an open-source modelling tool based on the Eclipse platform, Enterprise Architect, IBM Rhapsody or Catia Magic. Otherwise, there is software supporting dedicated MBSA frameworks as well. OPCAT, a patented software package, generates OPD and OPL for OPM (see Figure 54) (Dori 2011). CAPELLA is “*an Eclipse application implementing the ARCADIA (Voirin 2018) method providing both a domain-specific modelling language and a dedicated toolset*” (Roques 2016). XATIS (Safran 2022) is another software that supports CESAMES (Krob 2022) methodology for modelling complex integrated systems.

However, familiarising oneself with existing software supporting MBSA frameworks is often time-consuming and effort consuming. Some practitioners need help using the software even after proper training. This obstacle leads some practitioners to need still to work with conventional working software such as Excel, Word, PowerPoint, Visio, and CAD.

2.2.3 Discussion and Limits

Despite the lack of empirical evidence (Campo et al. 2023; Henderson et al. 2023; Henderson and Salado 2021), according to the value perceived by practitioners, the paradigm shift from a document-centric to a model-based approach improves, among other things, more completely capture architectural knowledge (Younse, Cameron, and Bradley 2021), better communication, information sharing, improved consistency, traceability, systems understanding, complexity management and capacity for data reuse (Campo et al. 2023; Henderson and Salado 2021). However, studies (McDermott et al. 2020; Chami and Bruel 2018; Cloutier 2015; Herzog, Hallonquist, and Naeser 2012; Maurandy et al. 2012; Andersson et al. 2009) also report that the complexity of MBSE methods and tools make adopting MBSE difficult and time-consuming. More solutions are needed to address the steep learning curve of architecture description framework and language, particularly for architects of complex technological systems who do not have a software engineering background. This thesis aims to alleviate these issues by proposing a solution that offers a more intuitive visual notation and a more efficient design process for all types of end-users.

2.3 Immersive Human-Computer Interaction

To improve the current situation, we decided to investigate the extensive Human-Computer Interaction (HCI) body of knowledge, especially immersive virtual environments, to strive for a more natural MBSA modelling interaction that natively embeds the fundamental concepts and

methodological aspects of the systems engineering approach without being contaminated by software engineering particularities.

2.3.1 Human-Computer Interaction

The term Human-Computer Interactions (HCI) began to draw attention with the book in 1983 (Card 1983). Dix (Dix 2003) depicts HCI from its fundamentals: the human user, the computer system and the nature of the interactive process.

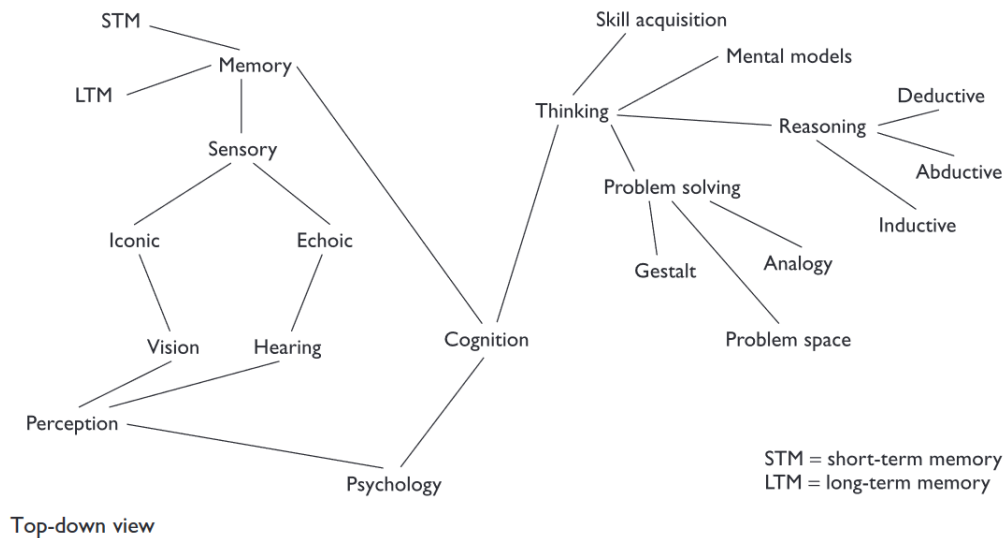


Figure 55 A top-down view of the semantic network of the human as an information processor (Dix 2003)

The human user is an information processor. It receives inputs from the world, stores, manipulates, uses information, and reacts to the information received. Figure 55 lists the elements of human perception and cognition. The computer system includes elements such as input devices and output display devices that affect the system user. The interaction involves translating what the user wants and what the system does. One popular interaction model is Norman's model (Norman 1988), which is as follows: 1) Establishing the goal, 2) Forming the intention, 3) Specifying the action sequence, 4) Executing the action, 5) Perceiving the system state, 6) Interpreting the system state, 7) Evaluating the system state concerning the goals and intentions. A framework providing basis-related interaction is shown in Figure 56. Norman's model (Norman 1988) and the framework (Dix 2003) contribute to analysing the interaction in terms of difficulty for a user to express his intention and recognise the state of the process.

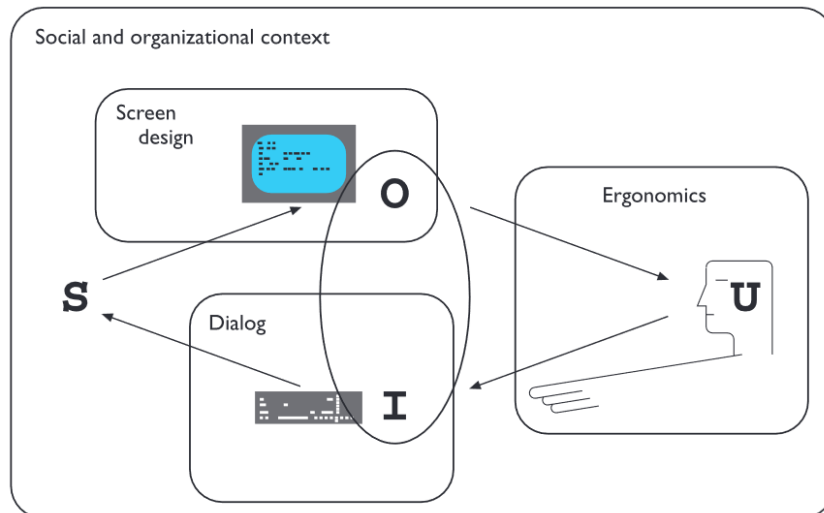


Figure 56 A framework for human-computer interaction (Dix 2003)

To sum up, HCI is a multidisciplinary field of study on the design of computer technology and the interaction between humans and computers.

2.3.2 Immersive Human-Computer Interaction

HCI focus on improving the efficiency and effectiveness of human-computer interactions. One approach to enhancing the HCI is using virtual reality technology, enabling access to 3D space to manipulate 3D objects. The concept of Virtual Reality (VR) was initially brought up in literary works before its renaissance in computer technology. It was first introduced by Sutherland in his PhD thesis (Ivan Edward Sutherland 1963) and invented with his other thesis (Sutherland 1965). Researchers began their studies on virtual reality since then. Depending on the study disciplines and the background of the researchers, they have their perspectives on defining virtual reality (Muhanna 2015):

- *“An immersive, interactive experience generated by a computer”* (Pimentel 2000).
- *“As any in which the user is effectively immersed in a responsive virtual world. This implies user dynamic control of viewpoint”* (F.P. Brooks 1999).
- *“Computer-generated simulations of three-dimensional objects or environments with seemingly real, direct, or physical user interaction”* (Dionisio, Iii, and Gilbert 2013).
- *“An integration of several elements, including computers, worlds and environments, interactivity, immersion, and users, usually referred to as participants in a virtual reality experience”* (Muhanna 2015).

The study's Virtuality Continuum (Milgram and Kishino 1994) defines the area between real and virtual environments as mixed reality (see Figure 57).

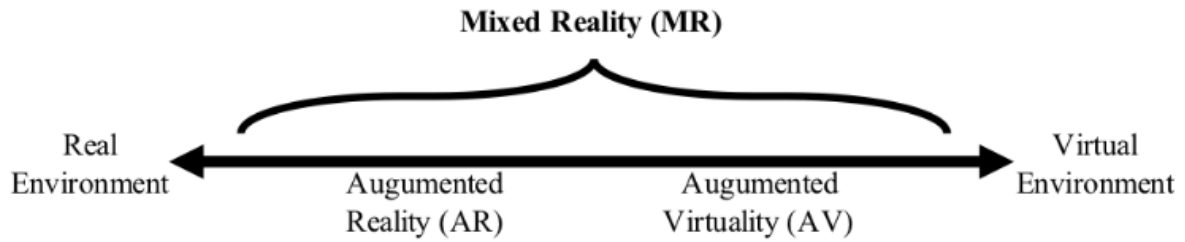


Figure 57 Simplified representation of a "virtuality continuum" (Milgram and Kishino 1994)

(Kommeter and Ebner 2019) proposed an adapted version of the classification of VR. (see Figure 58)

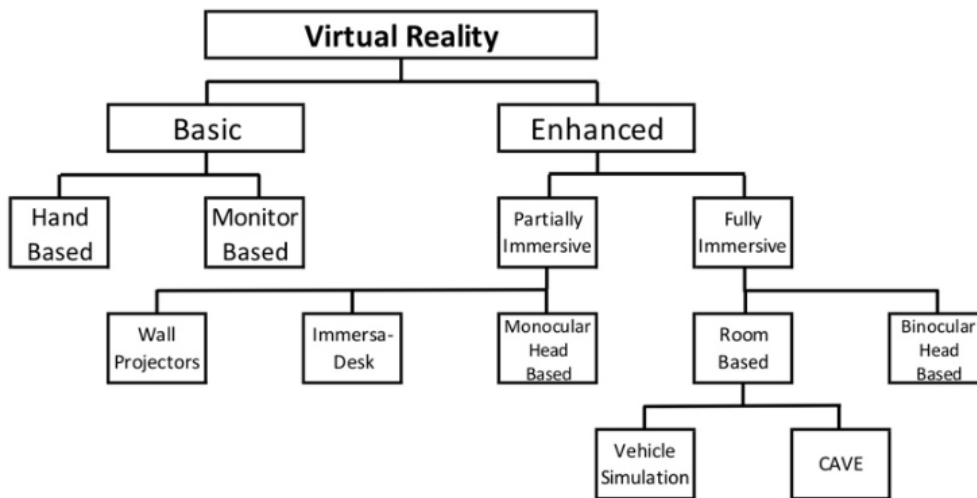


Figure 58 Classification of VR based on the type of technology used to achieve virtual worlds and interact with them (Kommeter and Ebner 2019)

To build practical virtual environment applications, it is crucial to optimise the basic interactions, in particular object manipulation, aiding users in concentrating on high-level tasks rather than low-level motor activities (Stanney 1995). Poupyrev classified the interaction techniques into selection, manipulation and navigation (Poupyrev and Ichikawa 1999). Bowman (Bowman, Johnson, and Hodges 2001) proposed a taxonomy for selection and manipulation techniques (see Figure 59). In another paper (Bowman, Koller, and Hodges 1997), Bowman proposed the taxonomy for the navigation techniques (see Figure 59).

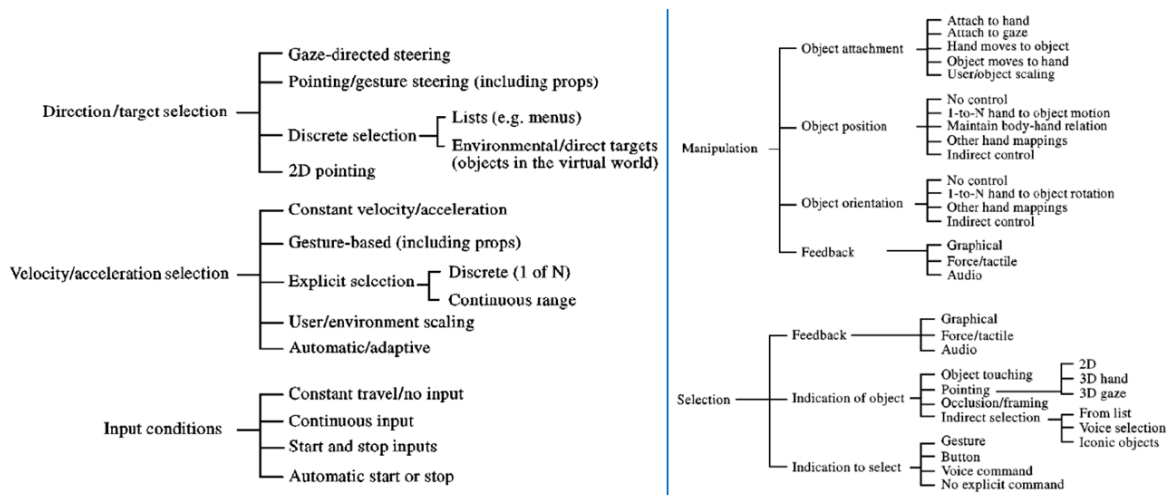


Figure 59 Navigation techniques (Bowman, Koller, and Hodges 1997) (left), manipulation and selection techniques (Bowman, Johnson, and Hodges 2001) (right)

Moreover, Figure 60 presents a well-established classification of VR manipulation techniques according to their primary interaction metaphors (Poupyrev 1998).

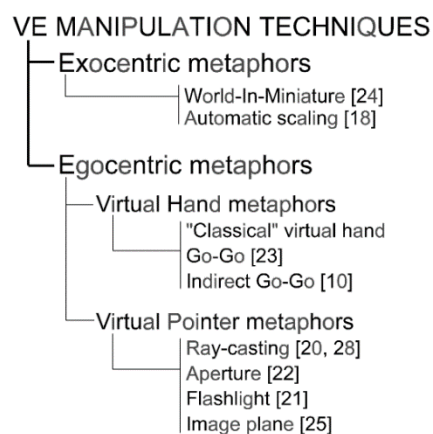


Figure 60 Classification of VE manipulation techniques depending on their underlying metaphors (Poupyrev 1998)

Based on the existing taxonomy, Raimbaud (Raimbaud 2020) proposed their taxonomy of characteristics for the VE navigation, selection, and manipulation techniques (see Figure 61).

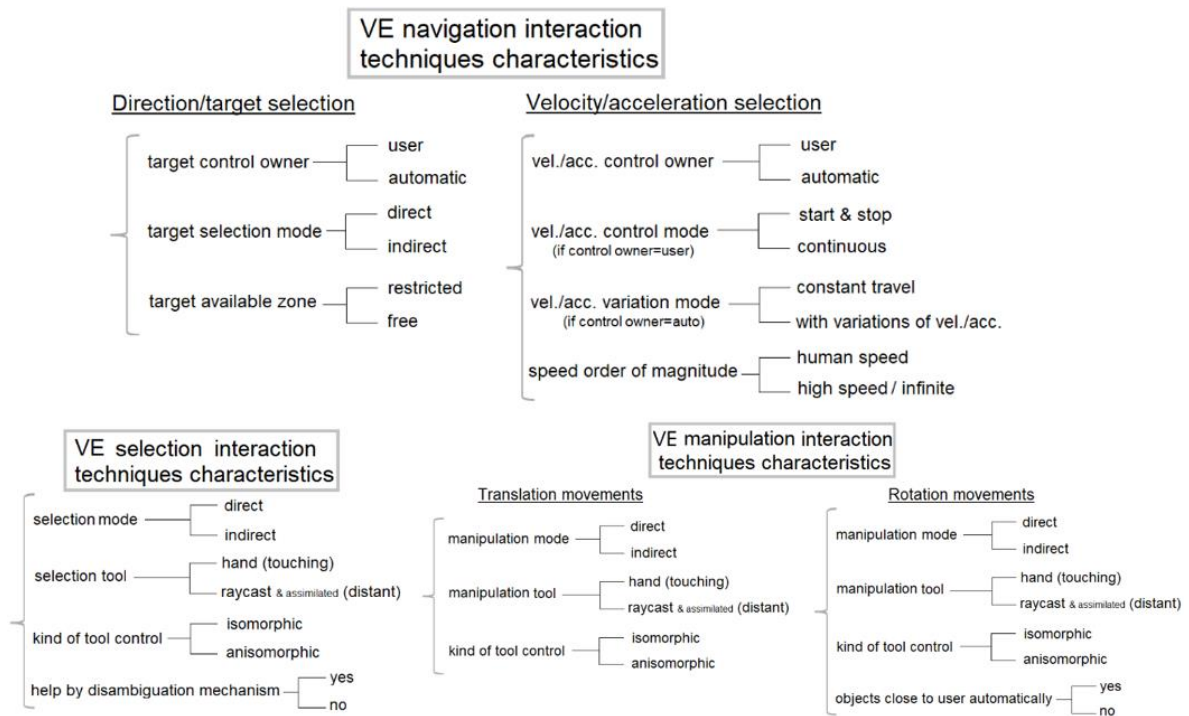


Figure 61 Taxonomy of characteristics for navigation, selection and manipulation (Rimbaud 2020)

2.3.3 Immersive Human-Computer Interaction for Software Architecting

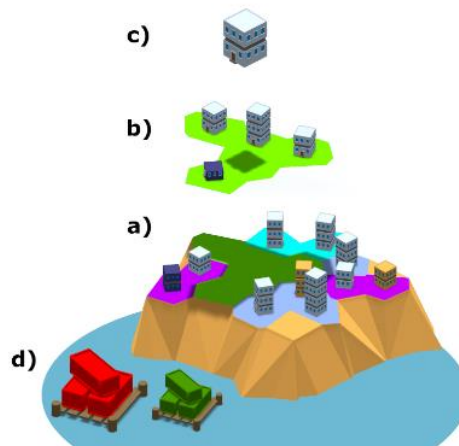


Figure 62 Island metaphor (Schreiber et al. 2019)

The advent of virtual reality helps us better understand complex projects with various abstract backgrounds. With the development of new cheap and mature Head-Mounted Displays (HMDs) between 2012 and 2016, VR became a feasible option for software tools to bring advantages like efficiency and effectiveness and, most importantly, enhance user experience in software development to conquer the invisibility of the abstract program code. Virtual reality in software visualisation systems drew attention in the 2000s. In (Schreiber et al.

2019), the island metaphor is used to visualise software architecture (see Figure 62). Similarly, there are other metaphors, such as the city metaphor (Vincur, Návrat, and Polasek 2017), molecule metaphor (V. L. Averbukh et al. 2004), cosmic metaphor (V. Averbukh et al. 2019), and geocentric metaphor (V. Averbukh et al. 2019).

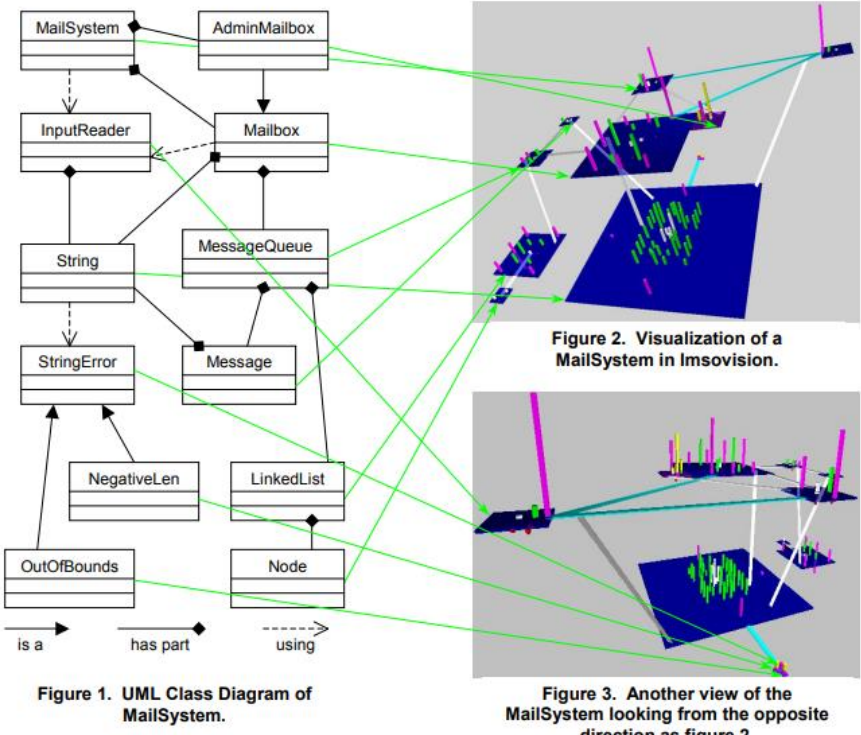


Figure 63 Example of visualisation (Maletic et al. 2001)

In 2001, Maletic proposed a visualisation metaphor for UML in an immersive virtual reality environment to assist users in comprehension tasks (Maletic et al. 2001). Figure 63 illustrates an example showing the related elements in different figures.

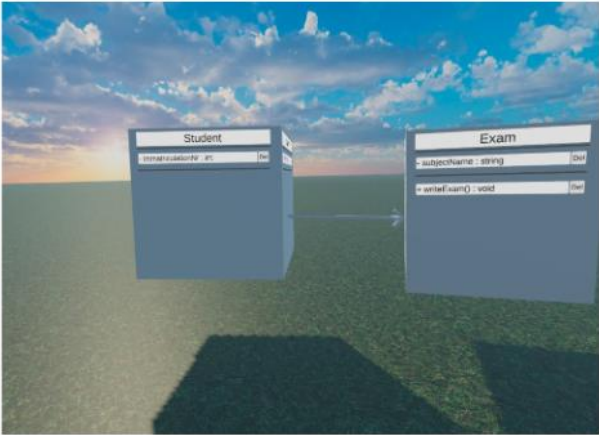


Figure 64 An UML Class Diagram example with custom 3D representation (Yigitbas et al. 2022)

Yigitbas proposed a collaborative UML modelling environment in virtual reality (Yigitbas et al. 2022) (see Figure 64) where they found out that “*the use of VR has some disadvantages concerning efficiency and effectiveness, but the user’s fun, the feeling of being in the same room with a remote collaborator, and the naturalness of collaboration were increased*” (Yigitbas et al. 2022).

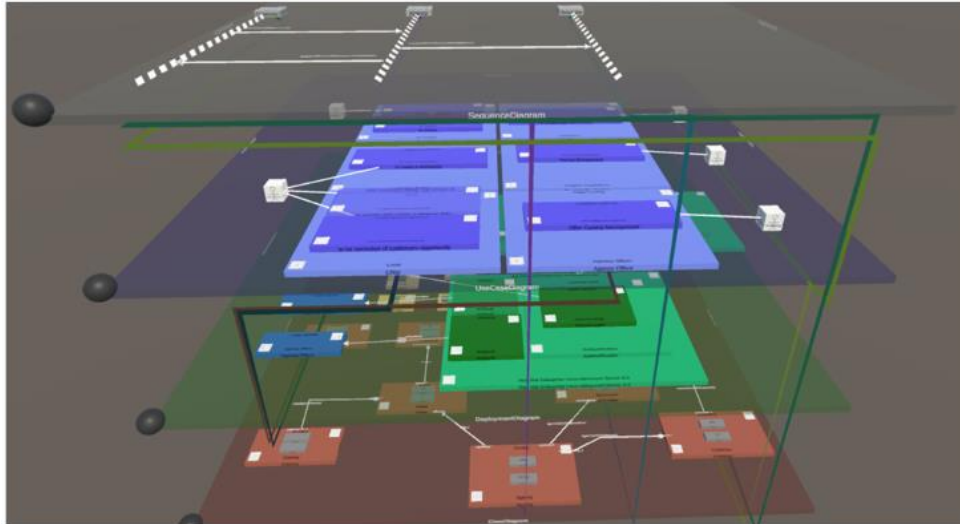


Figure 65 An example of VR-UML (Oberhauser 2021)

Regarding the use of UML in a virtual environment, another study (Oberhauser 2021) where Oberhauser proposed a slightly different metaphor (see Figure 65). Oberhauser also published studies about other languages in virtual environments. He refers to them as VR Modeling Framework (VR-MF). VR-MF “*provides a VR-based domain-independent hyper-modelling framework supporting multiple heterogeneous models while addressing three primary aspects of modelling in VR: visualisation, navigation, interaction, and data retrieval*” (Oberhauser 2021). Apart from VR-Unified Modelling Language (UML), it proposed VR-Business Process Model and Notation (BPMN) (Oberhauser, Pogolski, and Matic 2018), VR-Enterprise Architecture (EA) (Oberhauser and Pogolski 2019), and VR-Enterprise Architecture Tool (EAT) (Oberhauser, Sousa, and Michel 2020) as well. Figure 66 illustrates an overview of VR-MF.

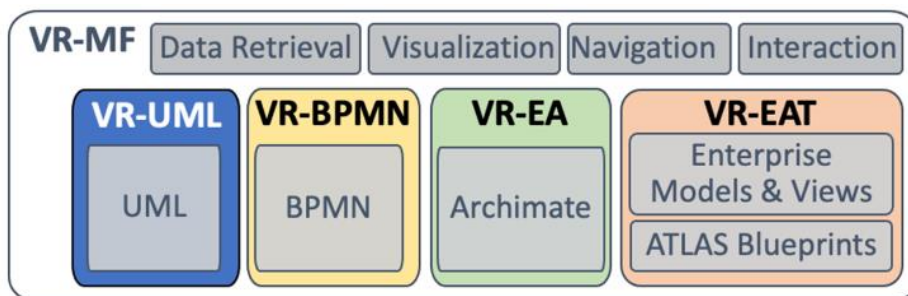


Figure 66 Overview of VR-MF (Oberhauser 2021)

2.3.4 Immersive Human-Computer Interaction for MBSA

In recent years, researchers have begun to use virtual reality for MBSE, focusing on MBSA. Some studies propose methods that allow users to review the system from one point of view, while others also allow editing. Sections 2.3.4.1 and 2.3.4.2 present studies on virtual reality for MBSA reading and authoring, respectively.

2.3.4.1 Immersive Human-Computer Interaction for MBSA Reading

(Mahboob et al. 2017) proposed a concept for the user and task-oriented model for product development in VR, focusing on the product context and using the MBSA approach. The proposed method in this paper tries to separate the scene into three parts: the product, the environment and the human actors. The context of the product in every lifecycle phase is different. It connects with the environment and the actors in different ways in each lifecycle phase. Thanks to this step, when a part of the scenario changes (often the case of the product during the evolution of development), they can minimise the resulting new tasks and maximise the reuse of achieved work. The project explores the construction of behaviour-containing scenarios and faster preparation of VR scenes.

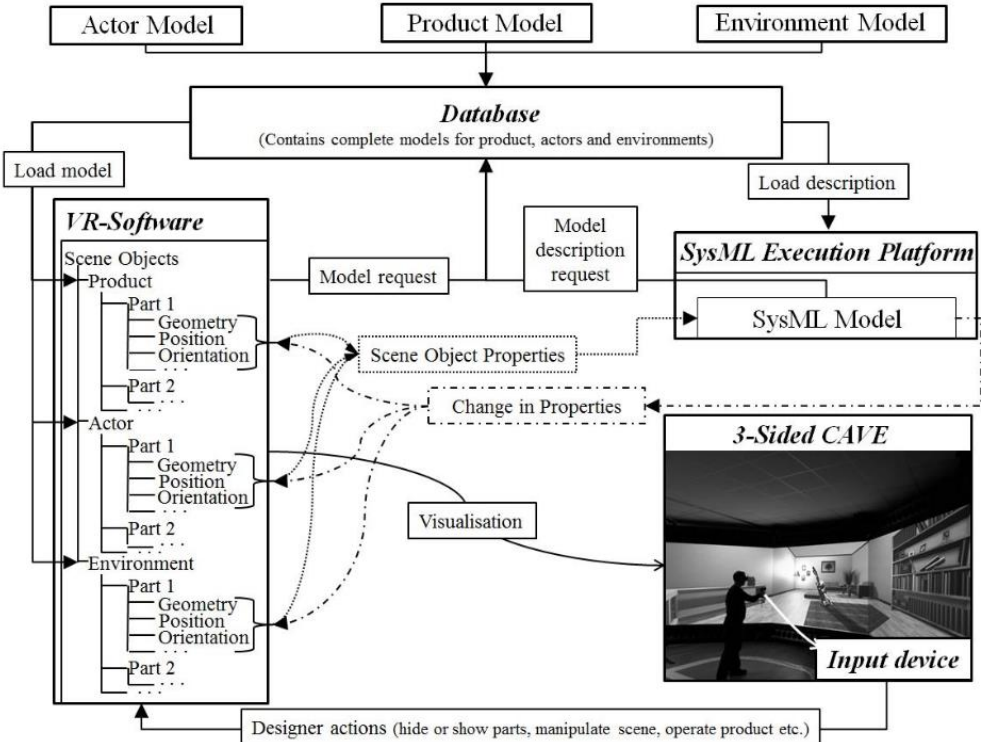


Figure 67 Overview of the complete process for VR scene configuration (Mahboob et al. 2017)

(Lutfi and Valerdi 2023) They have proposed an approach to visualising SysML diagrams in a virtual environment. They use a ground-based telescope system as the system's interest and apply the MBSE approach, developing it from the four pillars of SysML: requirements,

Behavior, Parametrics, and Structure. Afterwards, they export SysML diagrams into the virtual reality environment. Figure 68 demonstrates an example limited to the VR animation scenario of space objects (top) and an example of visualisation of a requirement diagram in a VR environment (bottom).

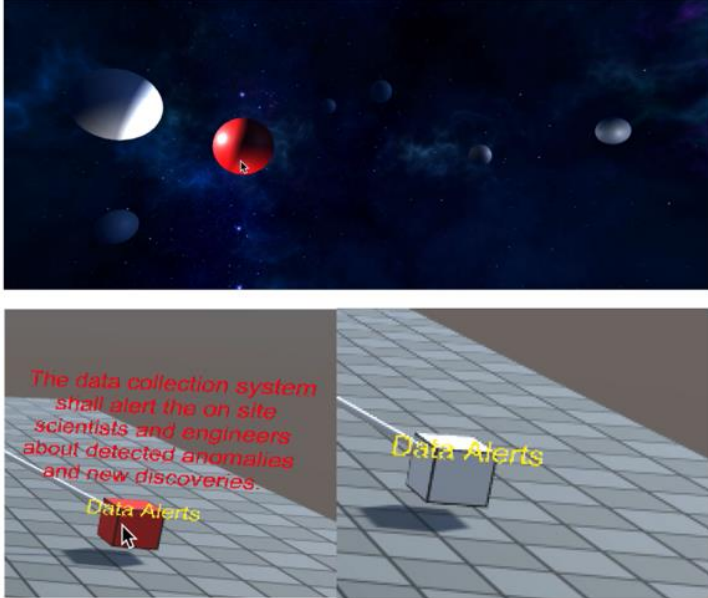


Figure 68 Examples of the visualisation method (Lutfi and Valerdi, 2023)

Another study by (Barosan and Van Der Heijden 2022) proposes to convert the data from the SysML models and visualise them in Unity 3D. The visualisation is through Unity 3D but not in virtual reality (i.e., the display device is not a VR device). The diagrams that can be extracted in the paper include a Block Definition Diagram, internal Block Diagram, State Machine Diagram, Use Case Diagram, and Requirement Diagram. An example of a filling station with a State Machine Diagram is presented in the paper (see Figure 69). The State Machine Diagram is represented on the top left of the screen.



Figure 69 3D visualisation of a filling station with a 3D State Machine Diagram (Barosan and Van Der Heijden 2022)

2.3.4.2 Immersive Human-Computer Interaction for MBSA Authoring

Regarding the authoring of data in a VR-based MBSA environment, the method (Korfatis, Cloutier, and Zigh 2012) present the system from the operational view to collect user requirements through the Concept Engineering System (CES). CES (see Figure 70) is the software System Engineering Research Center (SERC) developed to facilitate collaboration between stakeholders and developers in creating graphical and data models of stakeholder needs and proposed product concepts during the early stages of the development lifecycle. They believe the weakest link throughout the lifecycle of a system engineering is the connection between the user's needs and what the developers of the systems think the users need. The current system development environment calls for user needs to be specified in a Concept of Operations (CONOPS) (Korfatis, Cloutier, and Zigh 2012). It is demonstrated through an animation or movie to make it easier to represent the temporal dimension, as events occur in front of the viewer as a scenario unfolds. For the feedback they received from the workshop participants, there are pros and cons. The positive aspect is that the graphical representation and immersive environment lead to a much more realistic approach to CONOPS development, facilitating the observation of the anomalies and contradictions and enriching the visualisation of the scenario to make it more enjoyable. The detractor is that it is difficult to specify the links between objects and “intangible objects”, such as organisations.

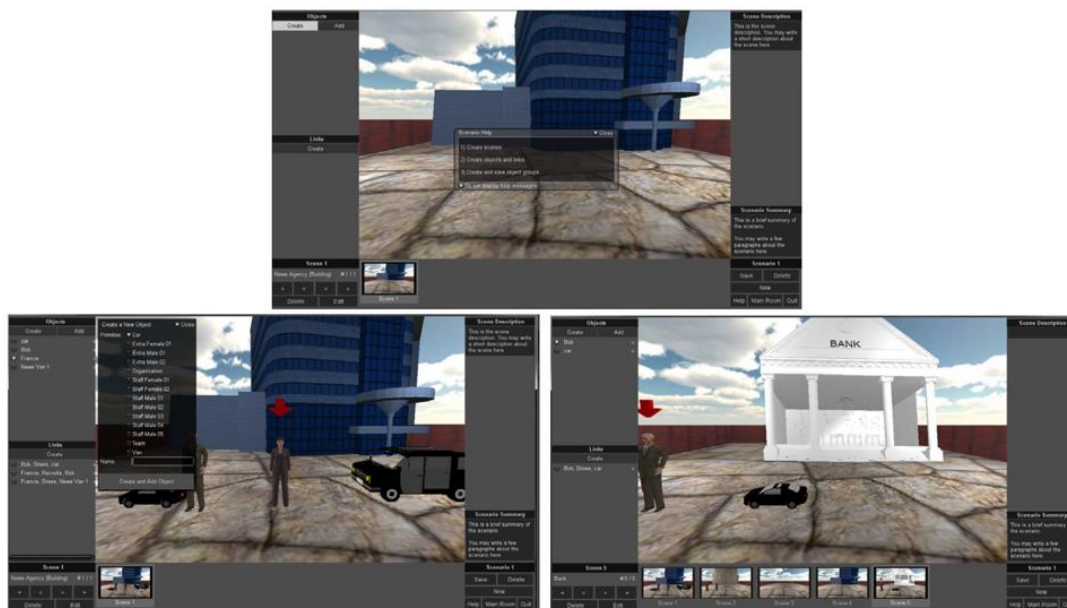


Figure 70 Overview of CONOPS: Screenshot of CES user interface (top) and two CES scenarios (bottom) (Korfatis, Cloutier, and Zigh 2012)

The same author who proposed the VR-MF (Oberhauser and Pogolski 2019) proposed VR-SysML in 2022 (Oberhauser 2022) as a new addition to VR-MF (see Figure 71).

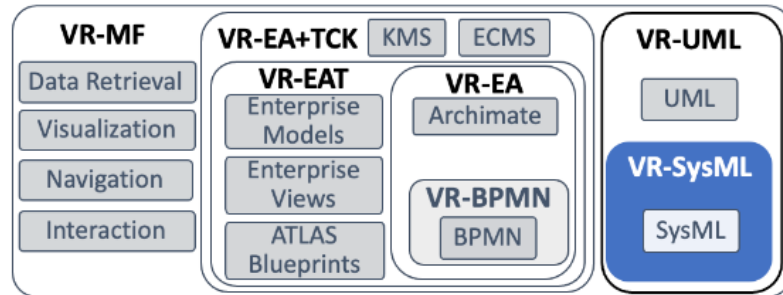


Figure 71 Conceptual map of various VR solution concepts in VR-MF (Oberhauser 2022)

VR-SysML visualises and edits the SysML models in the VR environment with an immersive experience. It automatically lays out views as stacked 3D hyperplanes, visualising the inter-view relations and element recurrence, and allows interactive modelling in VR (see Figure 72). Hypermodeling (i.e., display various models in the same space, provide automatic layout of views as stacked 3D hyperplanes, and visualise the reality of inter-view relations of elements) enables simultaneous visualisation of SysML in the same virtual environment, facilitating cross-model analysis across various diagram types and stakeholder concerns.

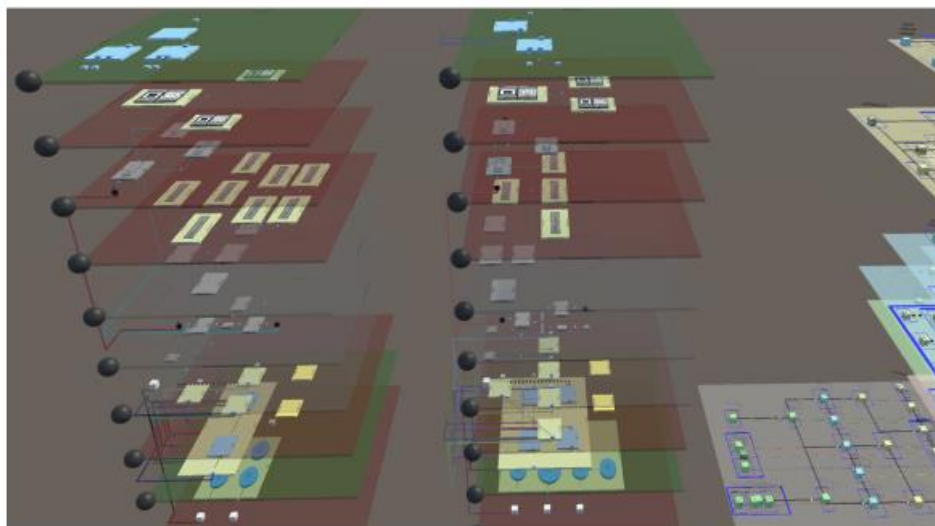


Figure 72 Multiple and heterogeneous side-by-side models in VR (Oberhauser 2022)

2.3.5 Discussion and Limits

Research on MBSA has been increasing recently, especially in terms of combining it with virtual reality. However, in these studies, we notice that most of the articles do not provide the evaluation part to valid their methods, such as in (Korfiatis, Cloutier, and Zigh 2012; Mahboob et al. 2017; Lutfi and Valerdi 2023) or just analytic evaluation for qualitative analysis, without end-user involved, such as in (Oberhauser 2022). Moreover, most studies do not propose any new visual metaphor for MBSA activities. Very few studies enable MBSA authoring in the

virtual environment. The study with novelty is the one in (Yigitbas et al. 2022), but their method is limited to one single class diagram in VR.

2.4 Validation of Immersive Human-Computer Interaction

As there are quality criteria and validation methods for conceptual modelling notations, immersive HCI has their own.

2.4.1 Validation Method of Human-Computer Interaction

The evaluation is not just a phase in the design process; it should be integrated throughout the design life cycle. The continuous feedback throughout the new HCI's design process guides the design's modification and improvements. It is not possible to do countless experimental tests all the time during the design phase. Instead, some informal evaluations would be constructive to get feedback on the current design. Meanwhile, some guidelines will be needed for such an evaluation approach. For instance, it can be some prototyping techniques, or it can also be the principles/theories we presented in Section 2.1.4. Those principles or theories can also help us during the design and evaluation phases.

One type of evaluation can be conducted with the designers or experts as expert analysis. The review assesses the performance of the design against cognitive and usability principles. Typically, this type of evaluation is analytic and qualitative, and it can take place throughout the design process or in the later design phase. Dix (Dix 2003) defines four approaches to expert analysis: cognitive walkthrough, heuristic evaluation, use of models, and use of previous work. Figure 73 summarises the four approaches of analytic evaluation techniques.

	Cognitive walkthrough	Heuristic evaluation	Review based	Model based
Stage	Throughout	Throughout	Design	Design
Style	Laboratory	Laboratory	Laboratory	Laboratory
Objective?	No	No	As source	No
Measure	Qualitative	Qualitative	As source	Qualitative
Information	Low level	High level	As source	Low level
Immediacy	N/A	N/A	As source	N/A
Intrusive?	No	No	No	No
Time	Medium	Low	Low-medium	Medium
Equipment	Low	Low	Low	Low
Expertise	High	Medium	Low	High

Figure 73 Overview of analytic evaluation techniques (Dix 2003)

Another type of evaluation is through user participation. Dix (Dix 2003) introduces the different approaches to assessment through user participation as: “empirical or experimental methods, observational methods, query techniques, and methods that use physiological

monitoring, such as eye tracking and measures of heart rate and skin conductance (Dix 2003)". Dix separates the evaluations by styles, such as laboratory and field studies.

- Experimental methods (see Figure 74): A controlled experiment is one of the most powerful design evaluation methods. It provides empirical evidence to support the research hypothesis. Moreover, *"it can be used to study a wide range of issues at different levels of detail"* (Dix 2003).

	Experiment	Interviews	Questionnaire
Stage	Throughout	Throughout	Throughout
Style	Laboratory	Lab/field	Lab/field
Objective?	Yes	No	No
Measure	Quantitative	Qualitative/ quantitative	Qualitative/ quantitative
Information	Low/high level	High level	High level
Immediacy	Yes	No	No
Intrusive?	Yes	No	No
Time	High	Low	Low
Equipment	Medium	Low	Low
Expertise	Medium	Low	Low

Figure 74 Overview of experimental and query evaluation techniques (Dix 2003)

- Observational methods (see Figure 75): Observation of users' interaction with the system is one of the popular ways to gather information for evaluation. Depending on the evaluation condition, they can be asked to complete a set of designed tasks in the laboratory or carry out their daily work in the fields. This observation can be conducted by think-aloud, protocol analysis, or post-task walkthrough.

	Think aloud ¹	Protocol analysis ²	Post-task walkthrough
Stage	Implementation	Implementation	Implementation
Style	Lab/field	Lab/field	Lab/field
Objective?	No	No	No
Measure	Qualitative	Qualitative	Qualitative
Information	High/low level	High/low level	High/low level
Immediacy	Yes	Yes	No
Intrusive?	Yes	Yes ³	No
Time	High	High	Medium
Equipment	Low	High	Low
Expertise	Medium	High	Medium

¹ Assuming a simple paper and pencil record

² Including video, audio and system recording

³ Except system logs

Figure 75 Overview of observational evaluation techniques (Dix 2003)

- Query techniques (see Figure 74): The query techniques help to get a view of the system from the user's perspective. Interviews and questionnaires are the two types of query techniques.

	Eye tracking	Physiological measurement
Stage	Implementation	Implementation
Style	Lab	Lab
Objective?	Yes	Yes
Measure	Quantitative	Quantitative
Information	Low level	Low level
Immediacy	Yes	Yes
Intrusive?	No ¹	Yes
Time	Medium/high	Medium/high
Equipment	High	High
Expertise	High	High

¹ If the equipment is not head mounted

Figure 76 Overview of monitoring evaluation techniques (Dix 2003)

- Evaluation through monitoring physiological responses (see Figure 76): One threat to the validity of the previous approaches is that the result depends on the researchers' observation or the user's subjective feedback. This can be eliminated by objective usability testing and by monitoring physiological aspects of computer use. Two popular ways are eye tracking and physiological measurement.

Furthermore, a methodology to design and evaluate a virtual environment (Gabbard, Hix, and Swan 1999) follows a sequence of four activities: user task analysis, expert guidelines-based evaluation, formative user-centred evaluation, and summative comparative evaluations. Figure 77 illustrates the proposed methodology. If we look at this thesis from the perspective of this methodology, user task analysis will be presented in Section 3. Expert guidelines-based evaluation will be presented in Section 4.6 with experts. Formative user-centred evaluation will be presented in Sections 4.4 and 4.5 with students. Summative comparative evaluation will be presented in Section 4.7.

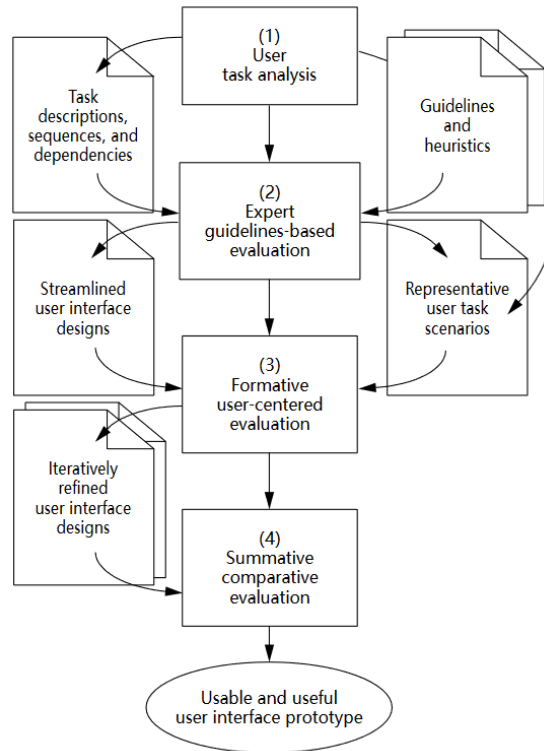


Figure 77 Methodology for the user-centred design and evaluation of VE user interaction (Gabbard, Hix, and Swan 1999)

2.4.2 Quality Criteria of Human-Computer Interaction

Section 2.1.4 presented several principles to serve as guidelines during the design phase or as evaluation criteria. Those principles are dedicated to the modelling language, while frameworks, languages, and software exist for an MBSA framework. Other criteria are required to broaden the scope of evaluation for a more comprehensive measurement of the proposed method.

(ISO 9241 2019) provides requirements and recommendations for human-centred design principles and activities throughout the life cycle of computer-based interactive systems. It provides the following principles for the human-centred design:

- The design is based upon an explicit understanding of users, tasks and environments.
- Users are involved throughout design and development.
- The design is driven and refined by user-centred evaluation.
- The process is iterative.
- The design addresses the whole user experience.
- The design team includes multidisciplinary skills and perspectives.

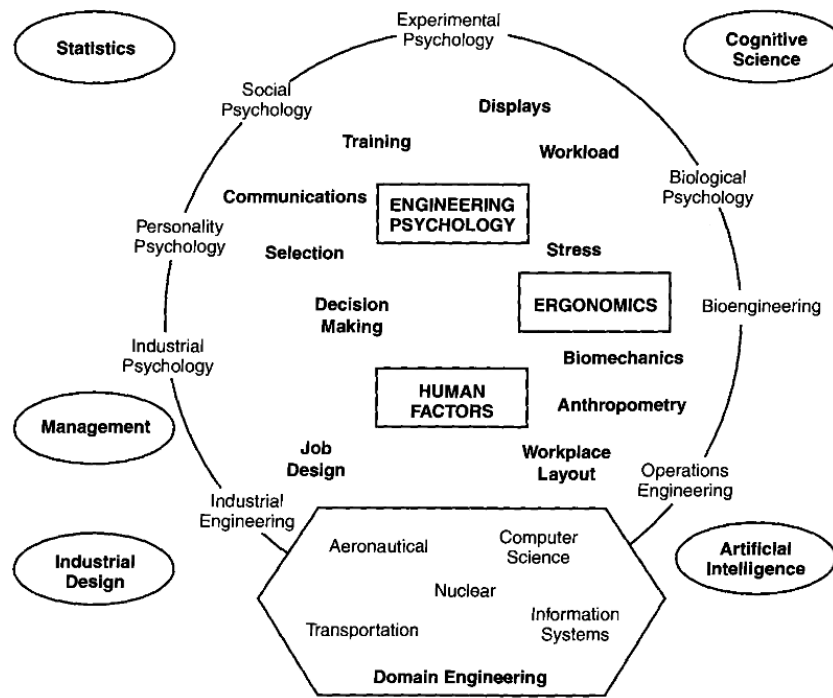


Figure 78 Relationship between human factors, shown at the centre, and other related disciplines of study (Wickens et al. 2004)

Ergonomics, or human factors, involves applying psychological and physiological principles to engineering and designing products, processes, and systems. The goal of ergonomics is to make the human interaction with systems one that reduces error, increases productivity, enhances safety, and enhances comfort (Wickens et al. 2004). (Wickens et al. 2004) defines the scope of ergonomics by illustrating the relationship of the discipline with other related domains of science and engineering (see Figure 78). The main concepts of ergonomics are in the centre of the circle, with subdomain studies surrounding it in bold. The subdomain studies within psychology studies are listed on the top, and subdomain studies within engineering studies are listed at the side. The disciplines outside the circle overlap some aspects of ergonomics.

Bach and Scapin (Bach and Scapin 2003) proposed an adapted version of ergonomic criteria for Human-Virtual environment interactions (HVEI). It includes eight elementary and 20 ergonomic criteria (see Figure 79).

1. Compatibility
2. Guidance
2.1. Legibility
2.2. Prompting
2.3. Grouping / distinguishing items
2.3.1. Grouping/ distinguishing by Location
2.3.2. Grouping/ distinguishing by Format
2.3.3. Grouping/ distinguish. by Behavior
2.4. Immediate Feed-Back
3. Explicit Control
3.1. Explicit User Actions
3.2. User Control
4. Significance of codes and behavior
5. Workload
5.1. Physical Workload
5.2. Brevity
5.2.1. Minimal Actions
5.2.2. Conciseness
5.3. Information Density
6. Adaptability
6.1. Users' Experience
6.2. Flexibility
7. Consistency
8. Error Management
8.1 Error Protection
8.2 Quality of error messages
8.3 Error Correction

Figure 79 List of Ergonomic Criteria for HVEIs (Bach and Scapin 2003)

Among the ergonomics, the Workload assessments were adopted for HCI research as “cognitive workload”, “cognitive load”, or “mental workload”. It describes the cognitive demands of a task. In this context, the NASA-TLX questionnaire (Hart and Staveland 1988) can be used to assess workload. NASA-TLX was not initially designed for cognitive workload evaluation in interaction with computers (HCI). However, it is the most popular questionnaire to measure cognitive workload (Kosch et al., 2023). NASA-TLX questionnaire (see Appendix H) splits the workload into six sub-scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration.

Usability is another factor that comes up often in evaluations. The definition of usability highly depends on the context since the term usability itself does not exist in any absolute sense. This indicates that there are no absolute measures of usability. ISO 9241-11 suggests that the scope of the measurement of usability should include measures of usability should cover:

- Effectiveness: The ability of users to complete tasks using the system and the quality of the output of those tasks.
- Efficiency: The level of resources consumed in performing tasks.
- Satisfaction: Users' subjective reactions to using the system.

An illustration (see Figure 80) from Di Gironimo et al. (2013) overviews the usability dimensions. It addresses the need for broad general measures of usability. Brooke (1996)

proposed a “quick and dirty” method that enables comparison across a range of contexts with low-cost assessments: the System Usability Scale (SUS). SUS is a ten-item 5 (or 7) point Likert scale questionnaire (see Appendix E.1), providing a global subjective view of usability.

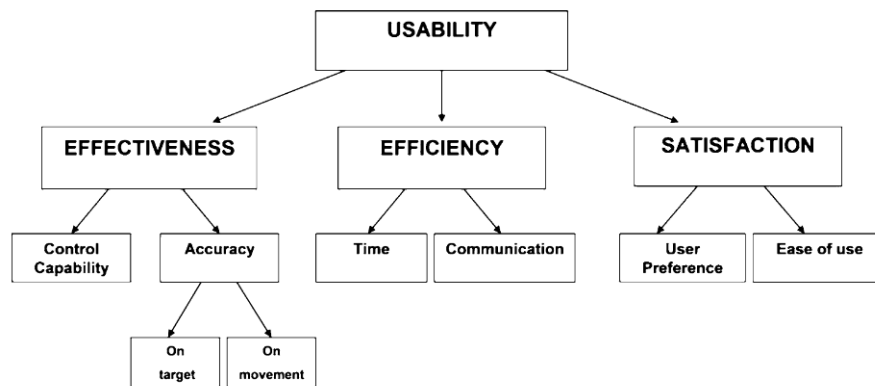


Figure 80 Overview of usability dimensions (Di Gironimo et al. 2013)

The study by (Bangor et al. 2009) evaluated the application of SUS with ten years of SUS application data collected on numerous products in all development phases. The results demonstrate that the SUS is highly robust and versatile as a tool for evaluations of usability. Furthermore, addressing the question “What constitutes an acceptable SUS score?” they proposed an acceptability score, quartile ranges, and adjective rating scale (see Figure 81).

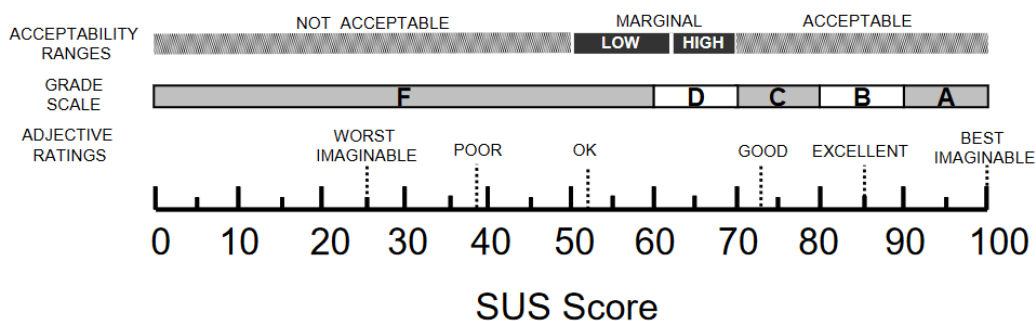


Figure 81 SUS score scale (Bangor et al. 2009)

For immersive HCI, Presence is one fundamental factor of VR, as the experience of “being” or “acting”. Assessment of Presence is typically done through questionnaires after the user’s experience of the VR environment. The study (Bangor et al. 2009) conducted a comparison among 15 published presence questionnaires where there are three popular questionnaires: IPQ (Regenbrecht and Schubert 2002), SUS(Slater and Steed 2000), and WS (Witmer and Singer 1998). They recommend the IPQ questionnaire since they find that IPQ provides the highest reliability within a reasonable timeframe. IPQ is constructed by four factors: one general item for General Presence, five items for Spatial Presence, four items for Involvement, and three items for Experienced Realism (Schubert, Friedmann, and Regenbrecht 2001). Therefore, the resulting IPQ has 14 items in total (see Appendix G).

2.4.3 Discussion and Limits

Among the papers involving HCI (either with the integration of virtual reality or not) in the MBSA, very few papers present the validation part of the proposed method. The claim is not backed by empirical evidence. It lacks a standard validation method for validating new HCI for the MBSE field. The papers that include an evaluation section always use indirect measurement. Direct physiological measurements are rarely used. The evaluation is often through observations from highly controlled tasks instead of realistic, complex tasks such as the ones performed by a system architect in this study. This leads to another problem as the evaluation is often done with the method designer or students rather than target users (e.g., architects).

2.5 Overall Discussion and Limits

This section presents the conceptual modelling notations, followed by the diagrammatic reasoning, which explains the cognitive process with diagrams. There needs to be more criteria to evaluate those notations from the perspective of visual syntax. The metaphor of graphic conceptual modelling languages is limited to conventional diagrams. This thesis proposed a new visual syntax, with the “grammar” of visual graphics included, based on studies exploring alternative visual graphics.

The terms and the standardised norms of MBSA activities are also presented with studies of immersive HCI applications. The MBSA approach aims to improve the system architecting activities. However, the adoption process in the field is complicated and time-consuming, particularly difficult for the novices. Furthermore, we presented some studies on MBSA's attempt to integrate virtual reality. Among the studies, only some provide empirical evidence for the validation, with most simply transferring representations into a virtual environment without substantial modification. Few studies allow users to engage in modelling activities directly in the virtual environment. Addressing those issues, the solution proposed in this thesis aims to provide a virtual environment where the user can engage in the MBSA activities with more intuitive interactions and visual graphics, improving the cognitive processes of the end-users for both experts and novices.

3 IMMERSIVE MODEL-BASED SYSTEMS ARCHITECTING INTERFACE

We propose a new immersive modelling interface to support MBSA tasks. To begin with the proposition, the MBSA tasks are identified first. The presentation of the virtual reality-based application starts with the definition of the data model and the modelling method. The application's architecture and the design of the visual and interactive metaphors will follow next.

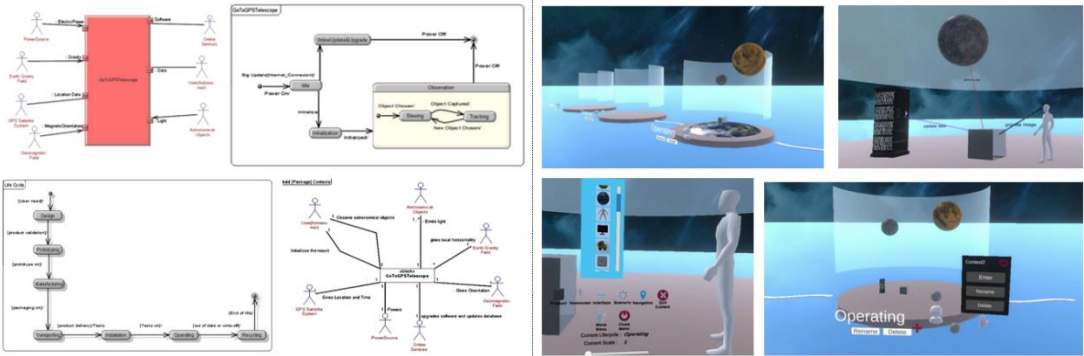


Figure 82 Illustrations of the 2D diagrammatic MBSA interface (Mhenni et al. 2014) and 3D immersive MBSA interface

3.1 MBSA Engineer Tasks Analysis

The introduction and literature review motivate research to invent new human-model interactions and, in a broader sense, new MBSA tools natively thought to meet the needs of systems architects and all people involved in developing a new engineered system. The development of new virtual environments requires designers to conduct a task analysis. Thus, before proposing new MBSA capabilities, the research method consisted of identifying and discussing the activities and tasks of a system architect before selecting the core tasks the new immersive MBSA architects will contribute.

The task analysis is based on the review of the systems engineering processes (ISO/IEC/IEEE 15288 2023) and model-based systems and software engineering processes (ISO/IEC/IEEE 24641 2023) ISO standards, as well as the existing MBSA methods.

3.1.1 Tasks from ISO 15288

ISO 15288 (ISO/IEC/IEEE 15288 2023) is an international standard that establishes a common framework of process descriptions for describing the life cycle of human-created systems. It defines a set of processes and associated terminology from an engineering viewpoint.

Business or Mission Analysis Process. This business or mission analysis process aims to define the overall strategic problem or opportunity, characterise the solution space, and determine potential solution class(es) that can address a problem or take advantage of an opportunity (detailed activities and tasks in Appendix B.1).

Stakeholder Needs and Requirements Definition Process. This stakeholder needs and requirements definition process aims to define the stakeholders' needs and requirements for a system that can provide the capabilities needed by users and other stakeholders in a defined environment (detailed activities and tasks in Appendix B.2).

System Requirements Definition Process. This system requirements definition process aims to transform the stakeholder, user-oriented view of desired capabilities into a technical view of a solution that meets the operational needs of the user (detailed activities and tasks in Appendix B.3).

System Architecture Definition Process. The system architecture definition process aims to generate system architecture alternatives, select one or more alternative(s) that address stakeholder concerns and system requirements, and express this in consistent views and models (detailed activities and tasks in Appendix B.4).

3.1.2 Tasks from ISO 24641

ISO 24641 (ISO/IEC/IEEE 24641 2023) is an international standard that deals with the tool capabilities and methods for model-based systems and software engineering (MBSSE). It specifies:

- a reference model for the overall structure and processes of MBSSE-specific processes and describes how the components of the reference model fit together;
- interrelationships between the components of the reference model;
- MBSSE-specific processes for model-based systems and software engineering; the processes are described in terms of purpose, inputs, outcomes and tasks;
- methods to support the defined tasks of each process;
- tool capabilities to automate or semi-automate tasks or methods.

The processes of MBSSE are categorised into four process groups: plan MBSSE, build models, perform MBSSE, and support models. Figure 83 shows an overview of the relationships among these process groups.

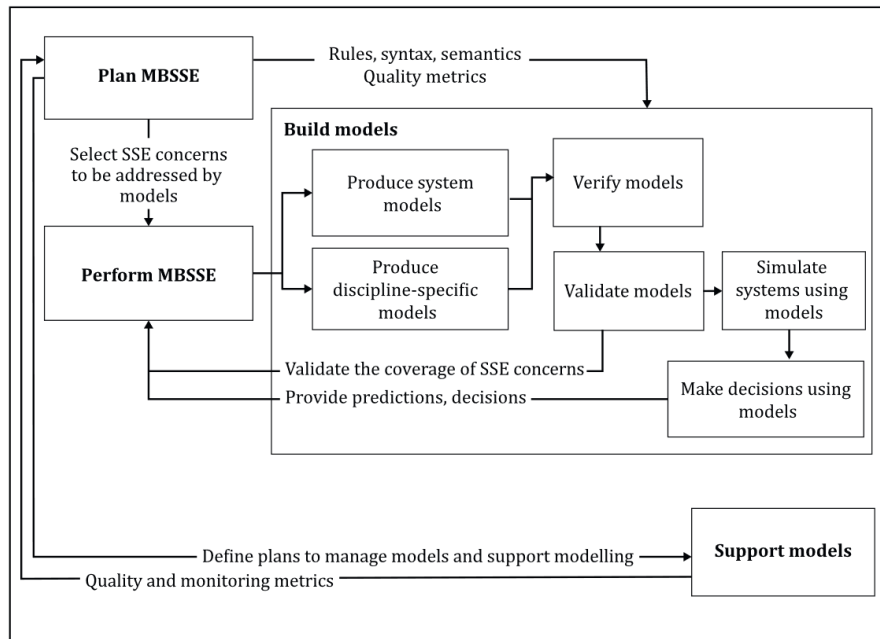


Figure 83 The relationships among MBSSE process groups and build models processes (ISO/IEC/IEEE 24641 2023)

The scope of this thesis is illustrated in Figure 84. Perform business and mission analysis, perform operational analysis, perform function analysis, perform system structure design, and produce system models will be presented here. Note that the wording of these processes is not exactly similar to those standardised in ISO15288 (ISO/IEC/IEEE 15288 2023) but are highly similar in terms of activities and tasks.

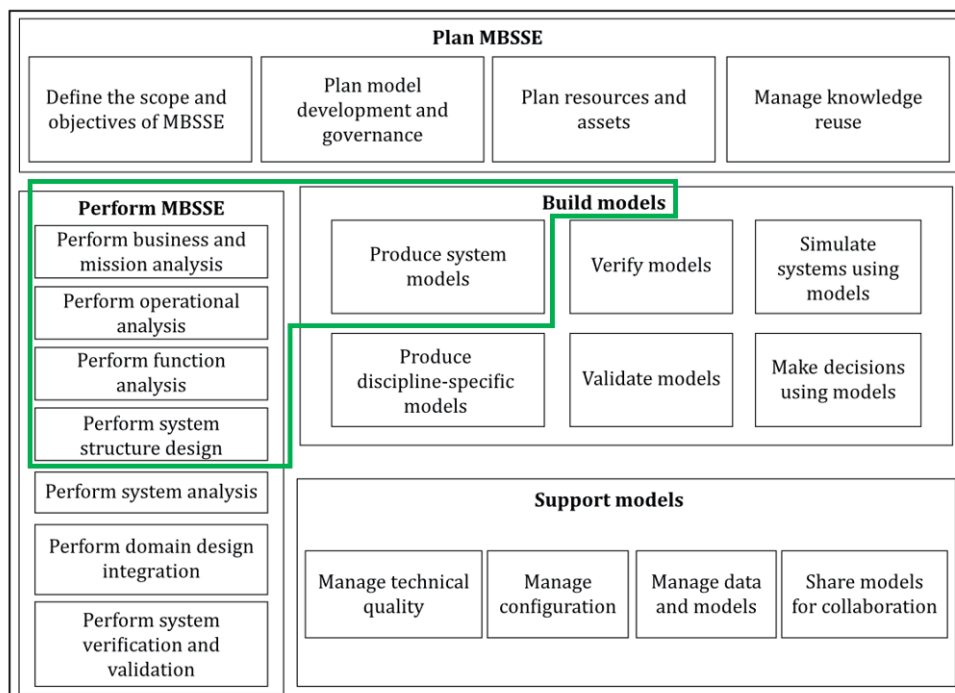


Figure 84 MBSSE reference model (ISO/IEC/IEEE 24641 2023)

Produce System Models Process. This process aims to produce systems models, either integral to the conduct of requirement and design engineering by domain experts or from engineering data collected from domain experts (detailed activities and tasks in Appendix C.1).

Perform Business and Mission Analysis Process. The purpose of this process is to establish the enterprise architecture model. One or more organisations can drive the enterprise, and the stakeholders can be from several organisations. The analysis of the business helps to identify the opportunities. The analysis of the enterprise mission allows the modelling of enterprise goals, definition of mission success criteria, modelling of mission environment, identifying the required capabilities, organisational structures and the solution overview (detailed activities and tasks in Appendix C.2).

Perform Operational Analysis Process. This process aims to define the OpsCon, which is initially prepared to support the concept and development stages of the system life cycle and then maintained throughout the program to support the production, utilisation, support, and retirement stages. Stakeholders and their needs differ according to the system context, which itself depends on the system life cycle stage. Models help to share a common understanding between stakeholders and to formalise the needs into Stakeholder requirements (detailed activities and tasks in Appendix C.3).

Perform Functional Analysis Process. This process defines the solution or system in terms of its functions, dysfunctions and performances. The model provides the basis for defining requirements and verification (detailed activities and tasks in Appendix C.4).

Perform System Structure Design Process. The purpose of this process is to structure the solution or system by grouping its functions. It includes the design of interfaces, physical structures and qualities. It provides data for speciality engineering analysis. This process is applied recursively at the element level, as elements are also considered systems. The model provides the basis for understanding how to describe an understanding of complex problems into a set of smaller and simpler ones while still ensuring the consistency of the whole (detailed activities and tasks in Appendix C.5).

3.1.3 Synthesis of Systems Architect Tasks from ISO 15288 and 24641

From ISO 15288, which is not limited to a model-based approach, the business or mission analysis process involves defining the problem and characterising the solution space and alternative solution space with the rationale of life cycle concepts. The stakeholder needs and requirements definition process involves defining the stakeholders and their needs and requirements for the system with the rationale on the context of the life cycle, stakeholders, the interface between the system and external entities, and the scenarios (or use cases). The system requirements definition process involves defining the system requirements and

functions with the rationale for the system requirements and functions. The system architecture definition process consists of determining the system architecture.

From ISO 24641, which prescribes a model-based approach, performing business and mission analysis involves defining high-level enterprise architecture, defining requirements, and generating the initial concept of operation (ConOps). Performing operational analysis involves identifying system lifecycle, boundary and context, identifying stakeholders, identifying use cases and scenarios, and capturing stakeholder requirements and MOEs. Performing functional analysis consists of identifying and decomposing system functions, identifying dysfunctions, developing functional flows and system states, identifying system requirements, constraints, and measures of performance (MOPs), tracing system operational activities to functions, and tracing stakeholder requirements to system requirements. Perform system structure design, which involves defining subsystems, defining functional interfaces, defining physical structures, and tracing logical system elements to physical elements. Producing system models consists of collecting the requirements of stakeholders, identifying the requirements of the system, identifying the interface between stakeholders and the system, identifying the system's logical structure or architecture, identifying system organic/physical structures or architecture, and analysing the system structure or architecture.

The processes from the two standards are different, but they share several recurring activities. Building a descriptive model for the system involves defining the lifecycle, context, stakeholders, stakeholders' requirements, use case (operational scenario), functions, interfaces, and system decomposition. Both standards also emphasise the importance of considering alternative solutions (or variant modelling) and maintaining product traceability across different modelling activities.

3.1.4 Tasks from Existing MBSA Methods

Apart from synthesising tasks from the standardised norms, the research on the literature review of the current MBSA practices to extract the concepts within the operational view is also conducted (see Table 6). This table shows the interpretation of the information captured in the operational view within SysML-based MBSE methods for modelling system architecture and the corresponding diagrams. The covered methods are ISE&PPOOA (Fernandez and Hernandez 2019), SYSMOD (Weilkiens 2020), HARMONY-SE (Hoffmann 2014), SE-READ (Mhenni et al. 2014), OOSEM (Friedenthal, Moore, and Steiner 2014; Walden et al. 2015), ASAP (Guillermo and Góngora 2014), SysCARS (J.-D. Piques and Andrianarison 2012).

Table 6 Review of the main SysML methods to identify the concepts and representations of an operational view

Method	Concepts of Ope. View	Diagram
ISE & PPOOA	Context Use and misuse cases Scenarios	Block Definition Use Case Activity
SYSMOD	Stakeholders Requirements Context Use case	Use case Requirement Block Definition Use case
HARMONY-SE	Services & stakeholders Use case context Use case scenarios	Use case Activity Sequence
SE-READ	Global mission Lifecycle Context External interfaces User operating modes Services Functional scenarios	Requirement State Machine Block Definition Internal Block State Machine Use Case Sequence
OOSEM	As-is operational domain Mission requirements To-be operational domain Mission scenarios Use case Context Operational states	Block definition Requirement Block definition Activity Use case Internal block State machine
ASAP	Mission Services Actors Needs	Use case Text Use case Requirement
SysCARS	Context Usage User scenarios Modes External interfaces Main services Systems scenarios States	Block Definition Use Case Sequence State Machine Internal Block Use case Sequence State Machine

3.1.5 MBSA Tasks Selection

The review of ISO 15288, ISO 2461, and existing MBSA methods (Table 6) shows no unique consensual ontology. Still, a set of common tasks requires defining MBSA concepts from mission definition to concept development. For the first version of the prototype, the list of tasks to define the MBSA concepts is not fixed or immutable but can be easily extended. The virtual environment for MBSA shall support the definition of:

- the system lifecycle,
- external entities that belong to a system context,

- external interfaces between the system and the external entities,
- services (or system functions) expected by the stakeholders,
- the operational (or use case) scenario,
- the subsystems,
- the internal interfaces between the subsystems,
- the recursive for each subsystem.

3.1.6 MBSA Data Model

Table 7 defines the MBSA concepts a system architect should define to achieve the selected tasks. These concepts' definitions are adapted from international standards or systems engineering guidelines, such as INCOSE guidelines. The concepts and their relationships are formally defined in the meta-model (see Figure 85), which underlies the 3D immersive MBSA interface.

Table 7 Concepts in 3D immersive MBSA interface

Design constraint	A limitation on the design or implementation of a system or a subsystem externally imposed by a stakeholder or external system.
External entity	A stakeholder or an external system.
External interface	A conceptual interface between the system-of-interest and an external entity that belongs to the system context.
External system	A system outside the system-of-interest but directly interacts with or indirectly influences the system-of-interest.
Function	An effect – intended by a stakeholder – of the interaction of the system with the system context.
Interface	An interface is a shared boundary between two (or more) systems or subsystems, defined by characteristics pertaining to functional or physical exchanges between them. It specifies the connection between two ports and, therefore, also the flow between the ports.
Internal interface	A conceptual interface between two subsystems.
Port	An interaction point and end of an interface.
Lifecycle	Series of all distinguishable lifecycle phases that a system goes through from conceptualisation until it ceases to exist.
Lifecycle phase	A period of time in the lifecycle of a system during which activities are performed that enable the achievement of objectives for that phase.
Functional Requirement	An agreed-to expectation for a system to perform a system function or satisfy a design constraint imposed on the system.
Scenario	At the system level, an (operational) scenario is a sequence of functions performed by the system-of-interest and external entities. At the subsystem level, a scenario is similar to a functional chain between external entities and subsystems.
Stakeholder	Individuals or organisations having a right to share, claim, or interest in a system of interest.
Subsystem	A combination of interacting elements that constitute a system and are organised to satisfy a set of subsystem requirements. The subsystem

	contributes to the realisation of higher-level system or subsystem requirements to which it belongs but does not, on its own, satisfy them.
System	A combination of interacting elements organised to satisfy one or more systems requirements.
System Context	System Context is the set stakeholders, external systems, and external interfaces surrounding the system of interest.
System-of-interest	The system whose life cycle is under consideration.

The selected MBSA tasks and concepts are well-established in the community. As defined in the meta-model, the lifecycle of the system consists of at least one lifecycle phase. In each life cycle phase, the system-of-interest is part of one or multiple system contexts. It interacts with external entities (systems or stakeholders) through external interfaces composed of two ports and one flow (multiple flows would require working out new visual metaphors so as not to overload the scene). By interacting with the external entities, the system-of-interest performs functions sequenced by at least one operational scenario, and the system requirements are derived from the system functional requirements. The same procedure can be recursively applied at the subsystem level (i.e., define the subsystems, their interfaces, functions, requirements, and functional scenarios).

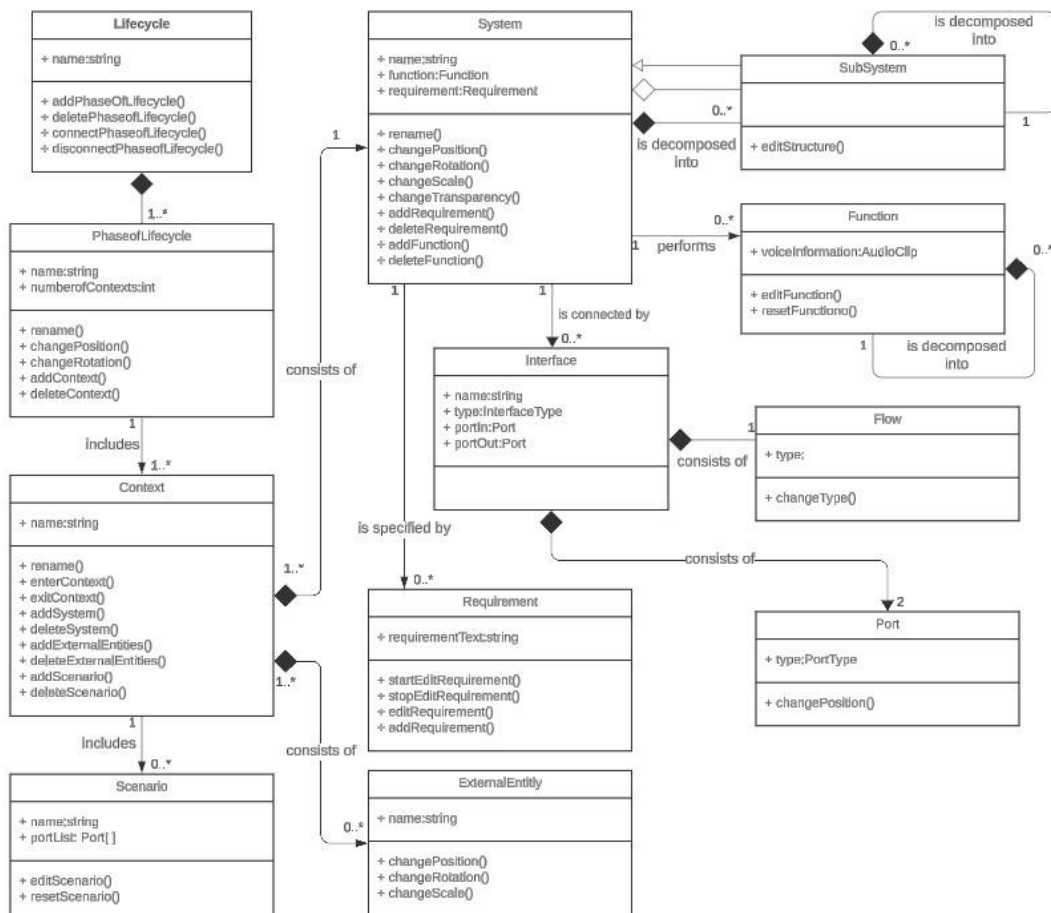


Figure 85 Data model of the 3D immersive MBSA interface.

3.1.7 MBSA Modelling Method

The modelling of these concepts requires the definition of a modelling method. The environment does not force a system architect to follow a specific step-by-step workflow; it remains relatively flexible. However, when we train practitioners who are willing to test the application, we encourage them to start with the definition of the system lifecycle. Then, they define one or multiple contexts for each phase of the system lifecycle. After creating a context, the user defines the external entities (stakeholders or systems) that belong to the context. The modelling of an external interface requires drawing a conceptual relationship between the system of interest and an external entity. Defining contexts for each system lifecycle phase helps to get a complete set of external entities and interfaces, limiting the risk of missing system requirements. After describing the system contexts, the system architect defines the system functions sequenced into operational scenarios. The architect derives the system requirements and design constraints from system functions and interactions between the system-of-interest and external entities. Finally, to satisfy the system requirements, the architect starts decomposing the system into subsystems, defining the internal interfaces between the subsystems before recursively performing the same MBSA activities. The sequence of activities is described in Figure 86.

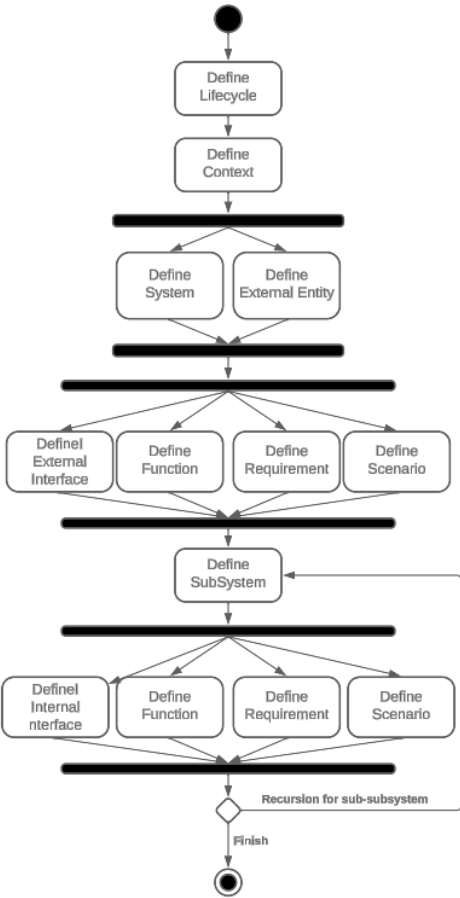


Figure 86 Overview of the method

3.2 Architecture

Figure 87 shows the architecture of the 3D immersive MBSA interface. The architecture comprises four parts: VR Environment, VR Devices, Visualization, and User.

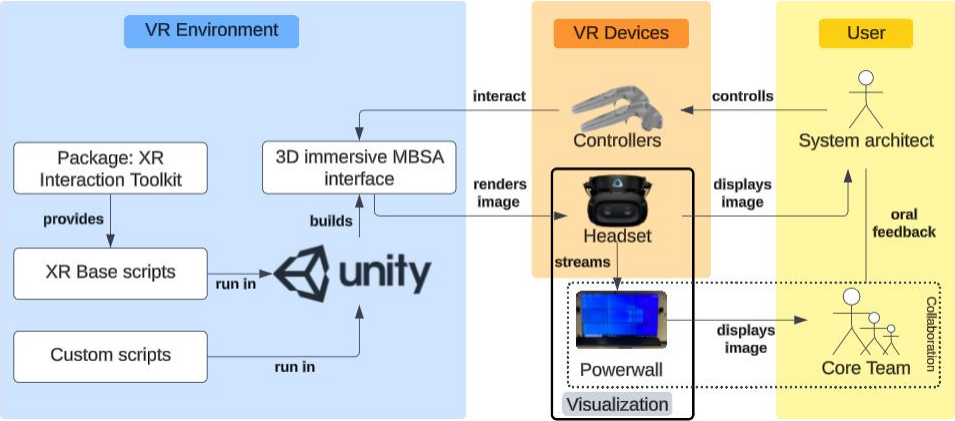


Figure 87 Overview of the architecture

The VR environment is built with Unity software, a powerful and versatile game development engine that allows developers to create immersive and interactive experiences across various platforms. It is widely used for developing video games, simulations, virtual reality (VR), augmented reality (AR), and other interactive content. In Unity, a package is a collection of assets, resources, scripts, and other files bundled together for easy distribution and reuse. In this thesis, we used several packages to build the environment. One of the core packages we used is the XR Interaction Toolkit. XR Interaction Toolkit is a Unity package designed to facilitate the development of VR and AR experiences. With the help of this package, the new 3D immersive MBSA interface is built in Unity.

The HTC Vive Cosmos VR headset interacts with the 3D immersive MBSA interface. The system architect can interact with the VR environment through the controllers and receive images from the headset. Other participants indirectly contribute to the architecture design in real-time via a Powerwall that streams the system architect’s field of view in the headset.

3.3 Visual Metaphors and Interactions

This section will present the visual metaphors of the semantic constructs used to encode the systems architecting concepts and explain how to interact with them. The 3D immersive MBSA interface is implemented based on the button actions from HTC Vive Cosmos. The instructions will use the buttons of this type of controller, as illustrated in Figure 88.

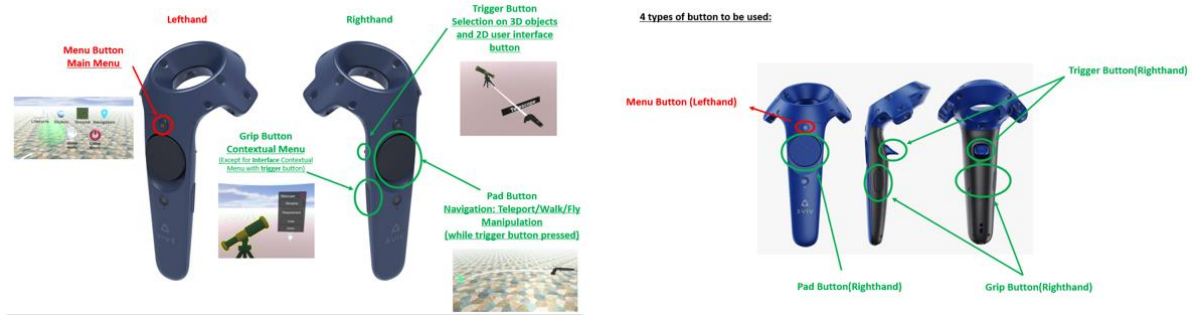


Figure 88 Illustration of buttons on the controller

Controllers: In total, the VR environment requires four buttons from the controller: the menu button of the left-hand controller, the trigger button, the grip button, and the pad button of the right-hand controller. The metaphor ray-casting is selected to interact with the VR environment (see Figure 89).

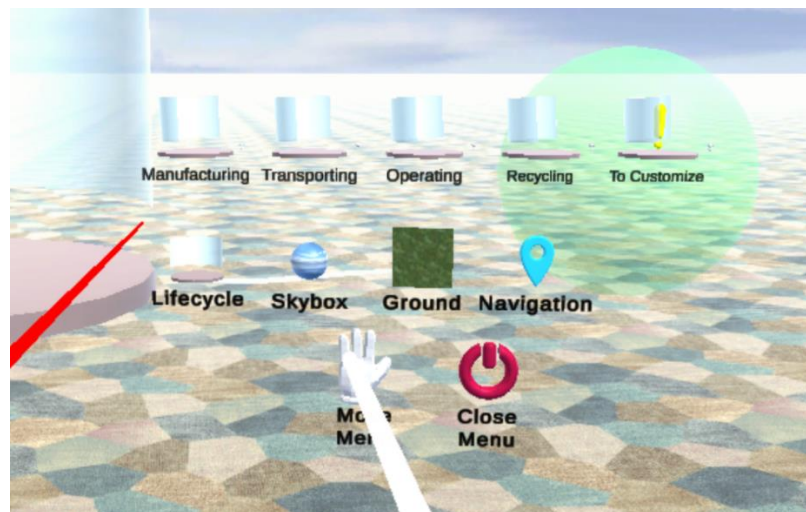


Figure 89 Illustration of ray-casting

Menus: In a 3D immersive MBSA interface, there are two types of menus: the main and contextual menus.

The main menu allows the user to access standard interaction options, such as the list to import 3D models in the environment or to change the texture of the virtual environment. It adapts depending on the current task being performed. Figure 90 shows the content of the main menu at the lifecycle and context definition tasks. To open the main menu, users must push the menu button on the lefthand controller.

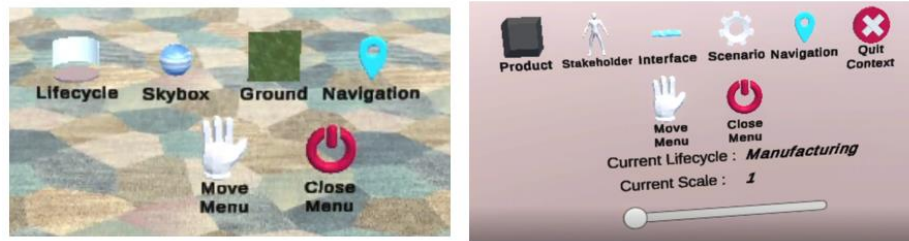


Figure 90 Main menu content at lifecycle definition tasks (left) and context definition task (right)

The contextual menu (see Figure 91) provides interaction options according to the selected object. For instance, the contextual menu of the external entities (here, a stakeholder) provides the option to modify its scale and name or access its requirements. The contextual menu of the system object enables users to break it down into subsystems. To open the contextual menu, the architect needs to point at the objects with the righthand controller and push the grip button on the righthand controller.

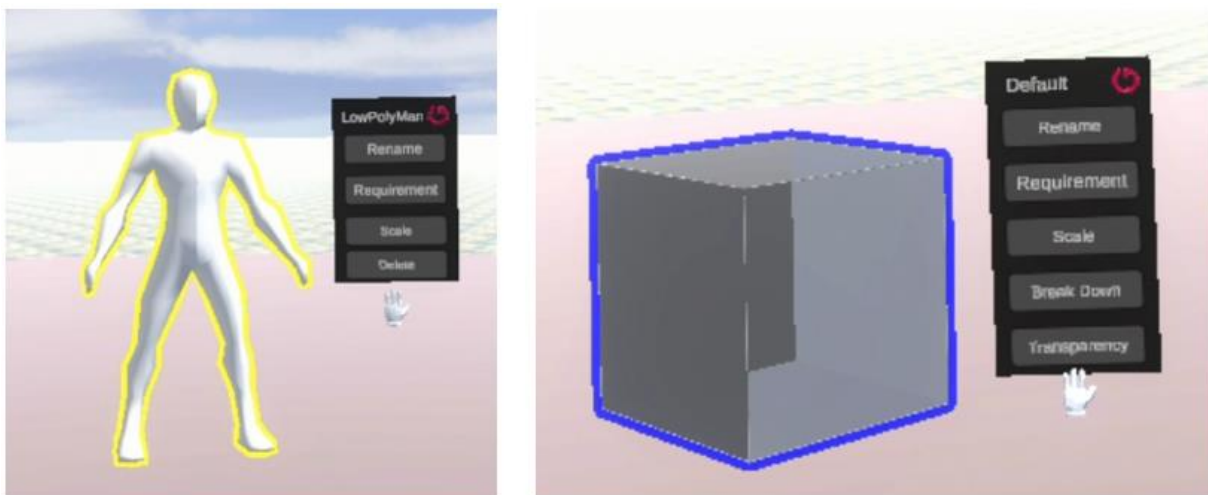


Figure 91 Contextual menu for a stakeholder (left) and the system of interest seen as a black box (right)

Navigation: The user can change the navigation mode from the main menu. There are three modes of navigation: Teleport, Walk, and Fly (see Figure 92). The default mode in the 3D immersive MBSA interface is teleport.



Figure 92 Navigation menu

Lifecycle: For encoding the concepts of a lifecycle, we designed an exocentric metaphor, that is, a God's eye viewpoint, where users interact with the virtual environment from the outside via the well-known World-In-Miniature technique (Poupyrev 1998). Each phase of the lifecycle is an empty World-In-Miniature (see Figure 93 right) that the user instantiates from a library by drag and drop anywhere in the 3D space. A 3D arrow connecting two World-In-Miniature metaphors encodes a transition between two phases. One phase can be attached to another by moving it to the transparent green sphere (see Figure 93 left). The translucent green sphere represents a socket for the attached phase of the cycle. A socket refers to a point or location on a 3D model where other objects can be attached or connected and where the position and orientation of the linked object can be preset. This auto-alignment will facilitate the manipulation and arrangement. Taking the example of the telescope case study, the lifecycle could consist of the following phases: designing, manufacturing, transporting, operating, and recycling.

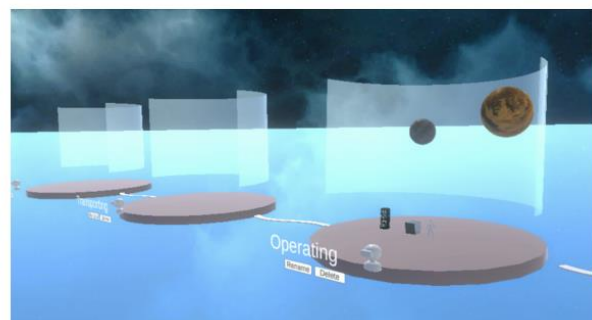
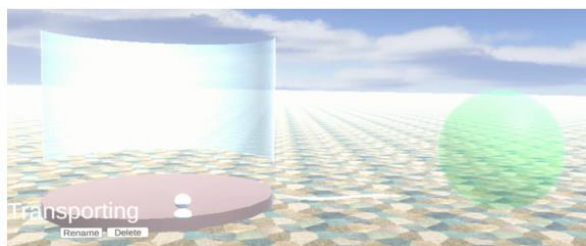


Figure 93 System lifecycle

Context: To create a context, the user can click on the icon (the little man wearing a head-mounted display) beside the name of the phase of the lifecycle, then the red plus (+) (see Figure 94). The grey sphere around the little man represents the context. The empty sphere is a better metaphor for the presentation to be discovered in the future, and it is used here since there has yet to be an apparent consensus on whether the 3D icon should present the context.

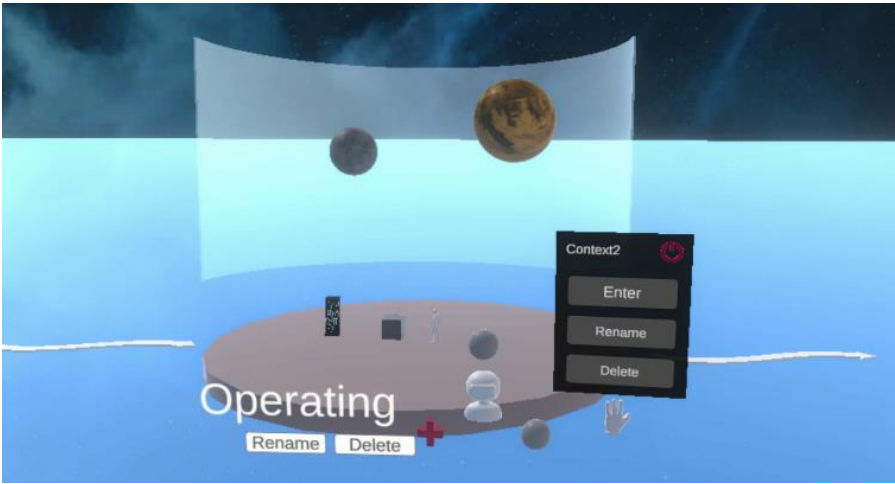


Figure 94 Example of creating a system context

To encode the context of a given phase of the lifecycle, we designed an egocentric metaphor in which the user interacts from inside the environment; that is, the virtual environment embeds the end-user's avatar (see Figure 95). Thus, when the user teleports inside a World-In-Miniature, he gets immersed in life-size.



Figure 95 System context

System-of-Interest: From the main menu, at the context editing level, the end-user can find the system library from which he can drag and drop 3D objects to instantiate anywhere in the context (see Figure 96). One of the common metaphors when we want to make an abstraction of any design solution, that is, any subsystem, is to see the system of interest as a

black box (from the external point of view of the system). Therefore, the black box metaphor represents the system in the context. The transparency and the 3D canonical shape of the black box are adjustable so that for the future development of the subsystem, the user can get an insight into the subsystem from the outside and choose the bounding box adapted to the system, respectively. Users can open the system contextual menu by pushing the grip button while pointing to the system, from which the architect can edit the name, transparency, scale, and requirement or go to the lower systemic level to define the subsystem.

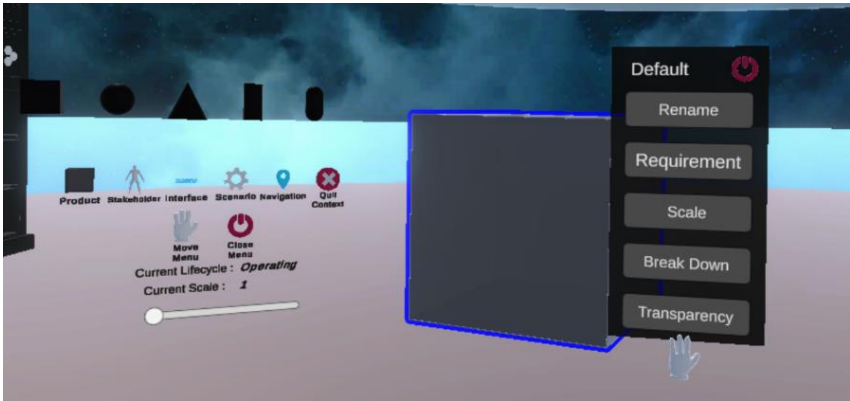


Figure 96 System with main menu (left) and contextual menu (right)

External Entity: Like a system, external entities can be instantiated from the library (see Figure 97). 3D models are used to present them. We use low—to medium-realistic models instead of high-level real models to avoid overloading the VR environment and maintain the conceptual design principles. Taking the example of the telescope case study, the external entities could be the server, user, and stars.

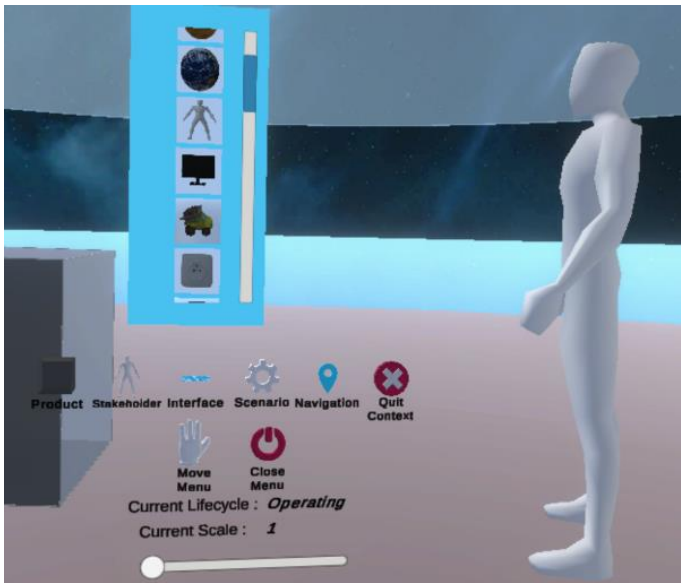


Figure 97 External entities library from the main menu

Requirement: The definition of requirements mainly relies on text input. Initially, we thought of using a VR keyboard for the input. However, with the current VR technology, the user experience with a VR keyboard still needs to improve. Instead of using a VR keyboard, we allow the end-user to switch from 3D to 2D. To edit a requirement, the user must stop the 3D stereoscopic mode to use the application in a 2D mode with conventional mouse-keyboard interactive devices. We believe there is no value in specifying text-based requirements in an immersive virtual environment. Moreover, it allows users to take a break from using VR devices. Indeed, with current technologies, spending more than 45 minutes in immersion takes much work. Finally, multiple stakeholders involved in the development of the system usually specify requirements. Our application could be accessible from desktop computers, and the scene's content can be reused to specify the requirements within the application or from popular requirements management software such as IBM Rational Doors. In the 2D mode, standardised templates (ISO/IEC/IEEE 29148 2018) of a requirement filled with the objects defined in a system context help the users to write good requirements. After editing, the user can go back to 3D mode and view the requirement list from the contextual menu of the system or external entity.

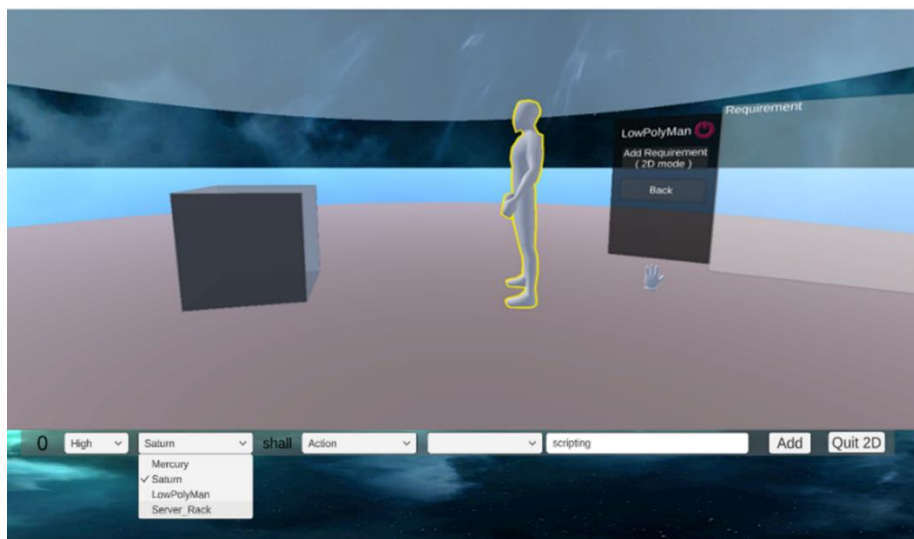


Figure 98 Requirement editing view on a 2D screen with the mouse and keyboard devices (top) and the requirement template (bottom)

Interface: The interface consists of two ports and one flow. A cube encodes the port concept, and a cylinder represents a flow. The current version has two types of ports: port in and port out. There are arrows on the faces of the cube to indicate the direction of the flow. Additionally, colours are used to differentiate the type of a port. Red is for port out, and blue is for port in. Users can enter the interface editing mode by using the interface menu from the main menu. While in the interface editing mode, the user needs to click on the system or an external entity to create the first port (port out) and then click on another external entity or the

system to build the second port (port in), finishing the creation of one interface. Once the user completes the definition of all interfaces, he can quit the interface editing mode from the menu before selecting the flow of an interface to open the interface menu, where the user can select the interface type. Taking the example of the telescope case study, the interface from the telescope (black box) to the server could be “Update data” with the type “Feedback”, the interface from the telescope to the user could be “Provide image” with the type “Human”, and the interface from the telescope to the star could be “Point to star” with the type “Control”.

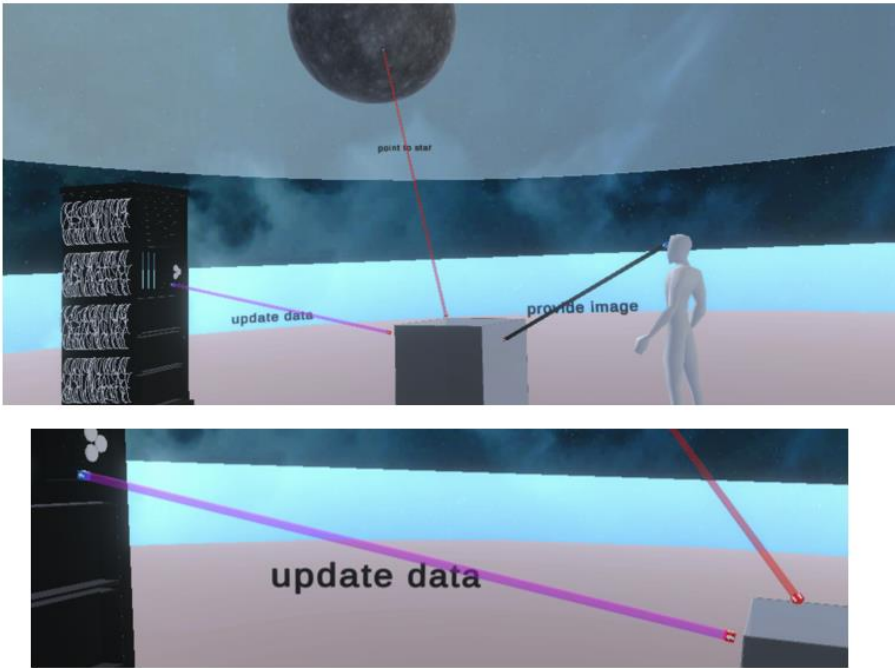


Figure 99 Interface with three flows (top) and a flow exchanged between an output red port and an input blue port (bottom). Textual statements attached to the flows correspond to the definition of functions.

The taxonomy of the interface types is borrowed from (Formentini et al. 2022). Figure 100 shows the table of Interfaces from the study.

Aircraft Interfaces				
Basic Flow	Type	Interface	Colour	RGB
Material	Human	N.A.	N.A.	N.A.
	Solid	N.A.	N.A.	N.A.
	Liquid	Fluid		230 230 250
	Gas	Air		7 255 62
	Human	N.A.	N.A.	N.A.
Signal	Acoustic	Electrical		255 255 0
	Chemical			
	Electrical			
	Hydraulic			
	Thermal			
	Pneumatic			
	Mechanical	Mechanical		138 43 226
Signal	Control	Electrical		255 255 0
	Feedback			

Figure 100 Interfaces (Formentini et al. 2022)

Function: As the conception of function is highly abstract and usually text-based, like the requirement, we created the function with another modality of input: verbal input. Verbal input can ease defining functions and allow users to perceive information from a different modality than the visual input. Functions are attached to the flows of interfaces (see Figure 101) for which one port is connected to the system-of-interest. It is inspired by existing diagrammatic modelling methods that use SysML or SysML-like sequence diagrams (Krob 2022) (see Figure 102). This modelling choice also helps to define the operational scenarios logically.

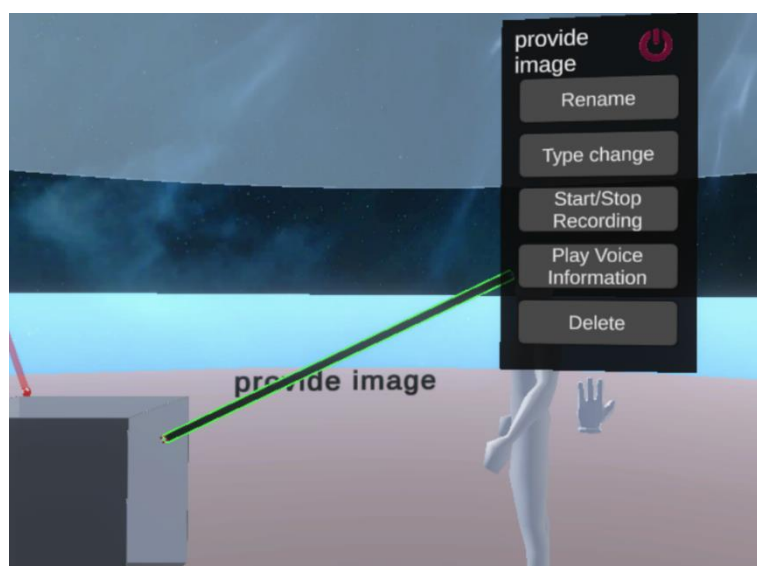


Figure 101 Function (play voice information button)

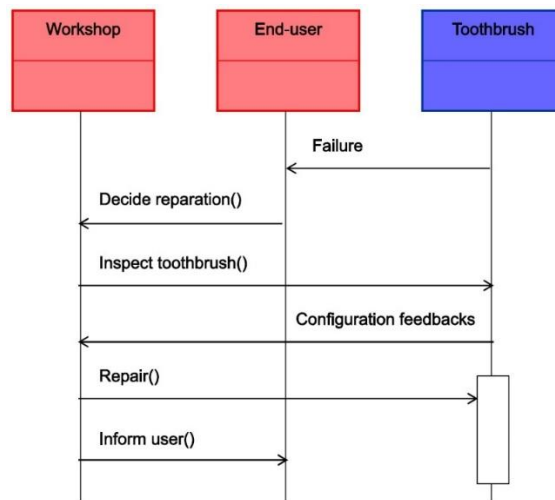


Figure 102 Example of operational scenario diagram for an electronic toothbrush (Krob 2022)

Scenario: The user can enter the scenario edit mode from the scenario menu in the main menu. To define a news scenario, the user selects a sequence of in/out ports of interfaces. The user should begin with the red (out) port and then select the corresponding blue (in) port, as it is the correct direction of the interface. The user should always finish the scenario edition by selecting a blue port. Thus, the user should select a sequence of red, blue, red, blue, [...], red, and blue ports for one scenario. When executed, the scenario is an animation that shows a green sphere navigating from one interface to another according to the selected sequence of interfaces during the edition (see Figure 103).

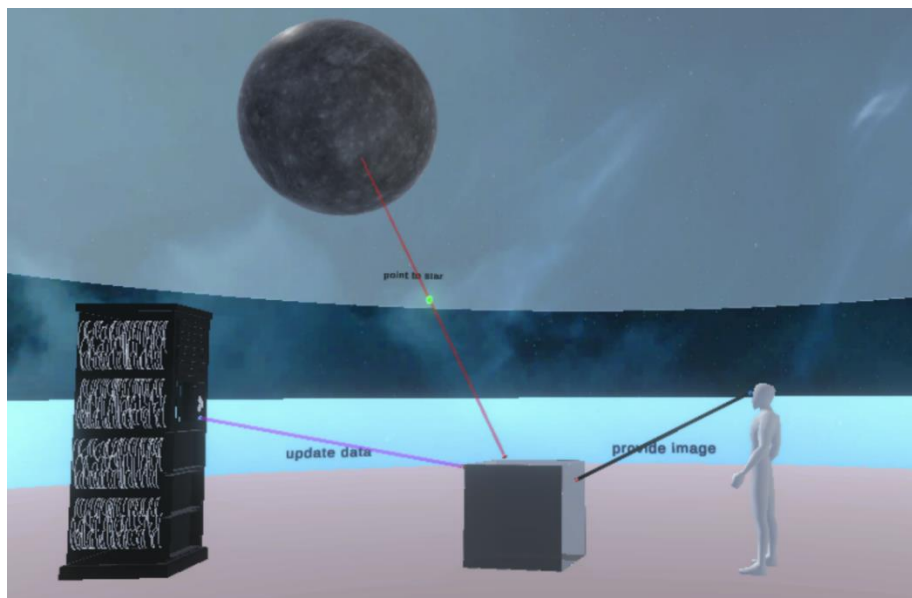


Figure 103 Scenario

Subsystem: The user can break the system down into subsystems. The user will immerse into the black box (system) and engage in subsystem modelling (see Figure 104).



Figure 104 Illustration from inside of black box

To model the subsystem, the user can either use primitives (sphere, cylinder, box, etc.) and perform Constructive Solid Geometry (CSG) operations (see Figure 105) or 3D models from the library (see Figure 108). CSG (Foley 1996) is a technique in solid modelling that enables designers to create complex objects using Boolean operations with canonical shapes. Figure 105 illustrates three Boolean operations on primitives.

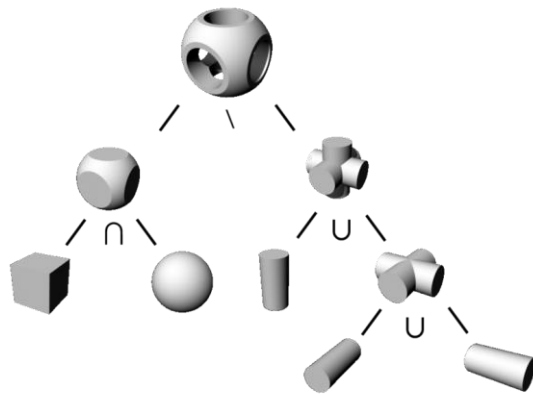


Figure 105 CSG Tree with the three Boolean operations: union \cup , intersection \cap , difference \setminus on primitives (Wassermann et al. 2016)

Users can perform the CSG operations in the 3D immersive MBSA interface (see Figure 106).

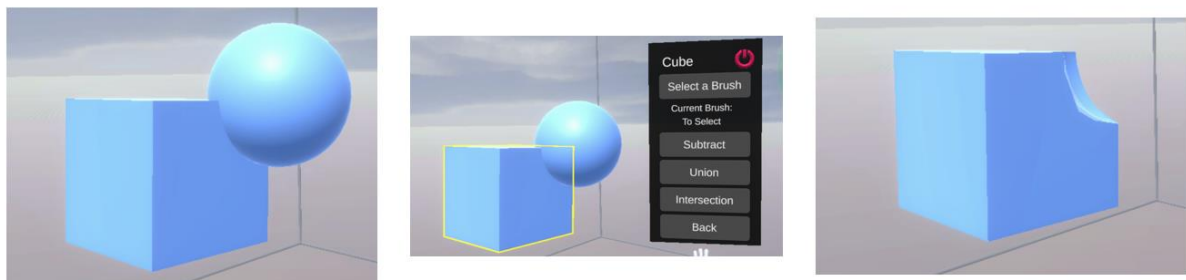


Figure 106 Example of subtract

Furthermore, a study compared solutions for 3D modelling in virtual reality, focusing on structure modelling (Tran, Foucault, and Pinquie 2021). The study proposes a VR-based environment to operate CAD B-Rep kernel editing operations with HMD and controller interfaces. In the 3D immersive MBSA interface, the user can also access a facet selection mode to select a specific primitive facet of a primitive 3D shape and modify it with direct push/pull mesh manipulation operations (see Figure 107 left). Finally, he can create 3D polygons, whether regular or irregular, by specifying the surface's vertices and determining the polygons' height (see Figure 107 right). However, in the current version of the prototype, conflicts arise among these functionalities. Still, it was not a problem since, due to limited experimental time, the validation will not require participants to evaluate these capabilities.

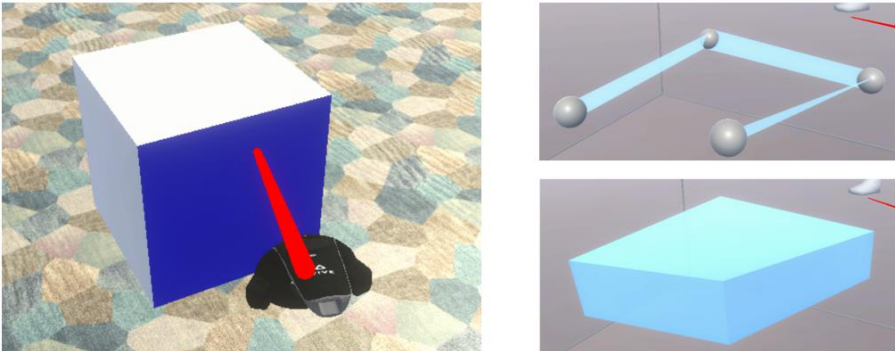


Figure 107 Illustration of facet selection (left) and polygon (right)

Once the subsystems are defined, the user can start the subsystem architecture definition as a recursion of the system analysis that includes defining the internal interface between the subsystems, their functions and requirements.

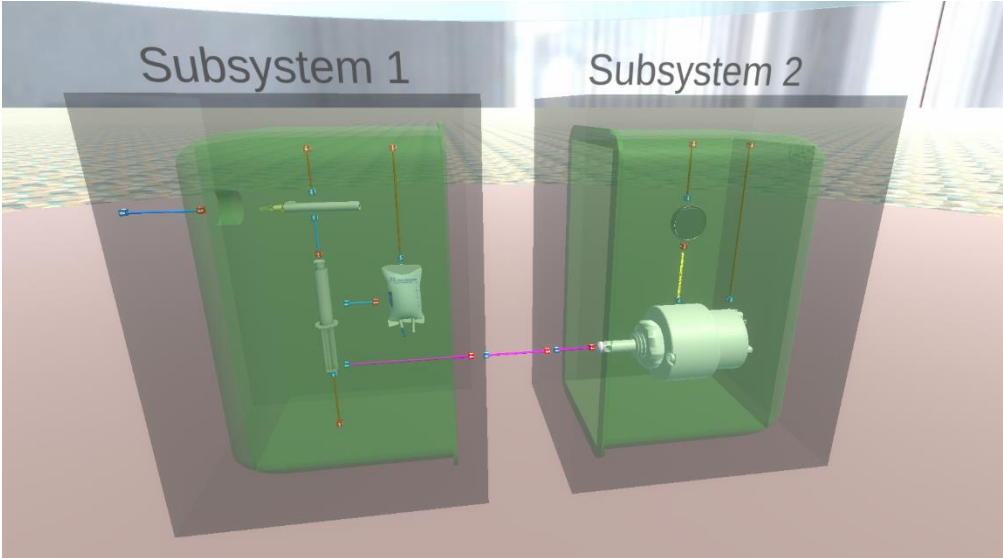


Figure 108 Subsystem

Free sketch: Users can draw free sketches using the 3D immersive MBSA interface (Grundel and Abulawi 2016). However, there needed to be more time for further research about efficiently managing this sketch element in the interface to help the design process.

Multiple users: The 3D immersive MBSA interface with the Netcode package in Unity can accommodate multiple users. However, due to the research's time constraints, only player avatars are synchronisable in this mode. Further implementation is required to synchronise the environment's objects and manage manipulation conflicts between two users.

Finally, Figure 109 presents MBSA models of another case study, a wearable on-body drug delivery system. The illustrations are extracted from the experimentation in section 4, where participants are asked to accomplish four tasks: the definition of the system life cycle (top left), its contexts of use (top right), the external interfaces (bottom left), and the operational scenarios (bottom right).

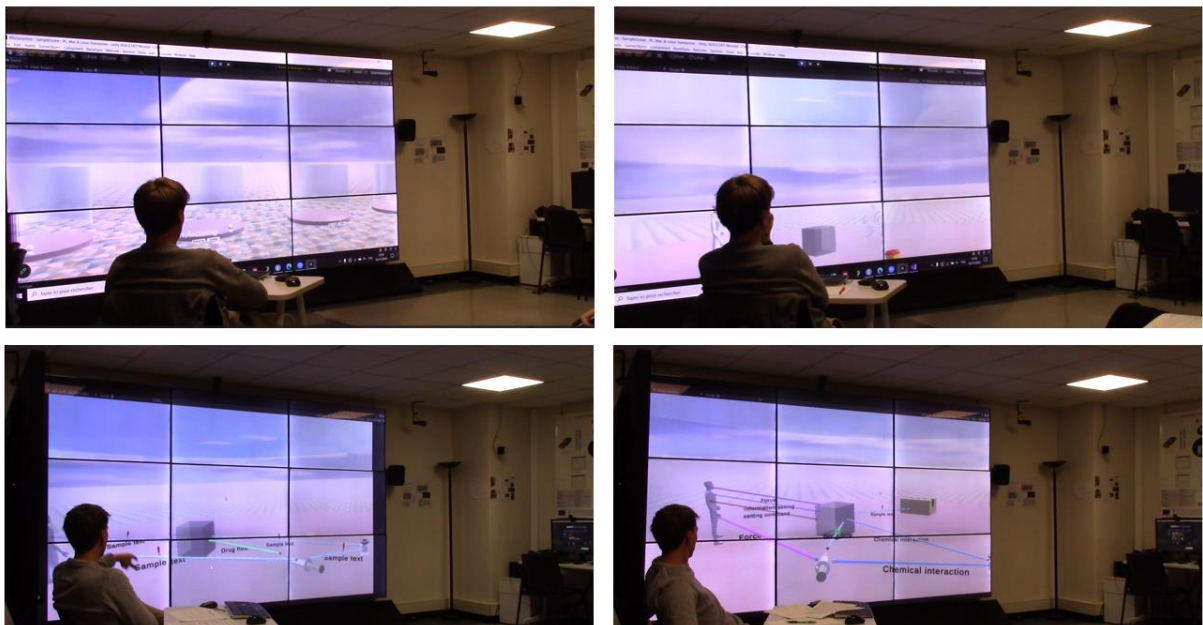


Figure 109 Photos of experimentation

4 EMPIRICAL VALIDATION

Despite the lack of empirical validation of research in engineering design (Barth, Caillaud, and Rose 2011; Isaksson et al. 2020; Vermaas 2014), this study was the subject of extensive work to provide scientific evidence that supports the proposition. All experiments were conducted with the functional prototype presented in the previous section.

4.1 Validation Method

This section details the validation strategy. Controlled laboratory experiments will be part of the validation process. Still, the validation of the proposal belongs to exploratory and formative research to gain insights, discover opportunities, and shape the directions for improving the current version rather than trying to find out what “the scientific truth” is. The methods used for data collection were highly influenced by several constraints, such as our capacity to recruit participants and their availabilities, the limited time to conduct the tests, etc. For instance, the involvement of domain experts is necessary to assess the usefulness and effectiveness or, in a broader sense, to inform the design of the 3D immersive MBSA interface. This constraint makes using controlled laboratory experiments to test our hypotheses difficult. Consequently, we will use a blend of research methods to evaluate the potential of using immersive 3D visuals to co-design a system architecture. This is of particular interest since the weight of evidence is strengthened when a combination of multiple – quantitative and qualitative – research methods are used to collect and analyse data. For each research question, the set of validation methods is summed up in Table 8.

Table 8 Validation methods and research questions

	RQ1	RQ2	RQ3	RQ4	RQ5
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

- Method 1: Online Survey (Survey)
- Method 2: Single-person controlled experiment (Expe_S)
- Method 3: Multi-person controlled experiment (Expe_M)
- Method 4: Formative usability tests (Formative)

Both single-person and multi-person controlled experiments with students (method 2 and method 3) will be used to compare the 3D immersive MBSA interface using an HMD with a classical 2D diagrammatic MBSA interface using conventional interactive devices (desktop, keyboard, mouse). The controlled laboratory experiments will be of interest to provide

statistical evidence. A post-experiment questionnaire serves to evaluate the validation criteria of the research questions.

We cannot recruit enough MBSA experts to participate in the controlled experiments, but their voices are more important than those of MBSA novices (i.e., the students in the university) as they are the true user-end group; meanwhile, the students are only potential users. Furthermore, they are more familiar with the MBSA adoption challenges than the students, and students are just aware of the systems architecting approach and related toolsets without any practical experience with real-world cases. Consequently, MBSA experts will participate in two formative usability tests: single-person sessions and multi-person sessions.

4.2 Selection of Quality Criteria for Evaluation of Immersive MBSA Interface

There is no design method to specify and develop an efficient and enjoyable immersive virtual environment that maximises user performance. Research studies that systematically investigate the human factors and design implications of immersive manipulation tasks, 3D devices, interaction metaphors and techniques need to be more extensive. Thus, designers have to rely on their intuition and common sense rather than on the guidance of established theory and research results in the practical development of virtual environments. Brooks (Frederick P. Brooks 1988) states: *“In watching many awful interfaces being designed (and in designing a few), I observe that the uninformed and untested intuition of the designer is almost always wrong”*. Rather than relying on our intuition, we specified, designed and conducted formative and summative evaluations of the MBSA virtual environment by leveraging a set of quality criteria that serve as heuristic guidelines. The quality criteria are directly related to the research questions.

User satisfaction (RQ1) will include the usability evaluation using the SUS questionnaire. Furthermore, similar to (Yigitbas et al. 2022), there will be additional questions on perceived efficiency, perceived effectiveness, retention rate, intention to use in the future, usage fun, and user motivation, as complementary to User satisfaction.

The assessment of the quality of the visual notation (RQ2) is based on the principles of the Physics of Notation (Moody 2009). We selected five principles out of 9 from the PoN, taking into account the constraints related to the online survey’s information delivery: Perceptual Discriminability (PD), Semantic Transparency (ST), Complexity Management (CM), Visual Expressiveness (VE), and Graphic Economy (GE). The excluded criteria are Cognitive Fit, Cognitive Integration, Dual Coding, and Semiotic Clarity. Cognitive fit refers to using different visual dialects for other tasks and audiences. This feature can be very interesting for future research, but for now, the interface only provides one set of representations of information.

Cognitive Integration consists of two mechanisms: conceptual integration and perceptual integration. Those mechanisms apply to the scenario when multiple diagrams present a system. This differs from the online survey, where the participants will only evaluate one diagram based on their first impressions. Therefore, this criterion is excluded from the questionnaire. The 3D immersive interface currently includes a few double-coding features. It is an interesting criterion as a cognitive fit for future work. Semiotic Clarity refers to Goodman's theory of symbols. (Goodman 1976) A notational system requires a one-to-one correspondence between characters and their referent concepts. When this requirement is violated, there can be one or more anomalies, such as symbol redundancy, symbol overload, symbol excess, and symbol deficit. This factor can be analysed directly by the language designer. Furthermore, the participants cannot assess this factor since they need to know all the elements (semantic constructs and visual syntax) in the language, which is impossible from just one diagram.

The presence (RQ3) and cognitive load (RQ4) are subjectively measured with the Igroup Presence Questionnaire (IPQ) and NASA-TLX questionnaire, respectively. Finally, perceived communication efficiency, communication effectiveness, and member activeness are measures of group dynamics (RQ5).

Table 9 Overview table of criteria

RQ1: User satisfaction	
Usability (SUS)	A questionnaire on evaluation of usability covering effectiveness, efficiency, and user satisfaction.
Perceived efficiency	User's perceived level of resource consumed in performing tasks.
Perceived effectiveness	User's perceived ability to complete tasks using the system and the perceived quality of the output of those tasks.
Perceived retention rate	User's perceived ability to reproduce the presented information.
Intention to use in the future	The user intends to decide whether to use the system in the future.
Usage fun	User's fun while using the system.
User Motivation	User's motivation while using the system.
RQ2: Visual Notation (PoN)	
Perceptual Discriminability	The ease and accuracy with which symbols can be differentiated from each other.
Semantic Transparency	The extent to which the meaning of a symbol can be inferred from its appearance.
Complexity Management	The ability to represent information without overloading the human mind.
Visual Expressiveness	The number of visual variables used in a notation.

Graphic Economy	Limit the number of graphical symbols in a notation to consciously maintain the meanings of symbols in working memory.
RQ3: Presence (Igroup Presence Questionnaire)	
Presence	The general “sense of being there”.
Spatial Presence	The sense of being physically present in the VE.
Involvement	We are measuring the attention devoted to the VE and the involvement experienced.
Experienced Realism	We are measuring the subjective experience of realism in the VE.
RQ4: Cognitive Load (NASA-TLX Questionnaire)	
Mental Demand	The cognitive effort required to complete the task.
Physical Demand	The physical effort required to complete the task.
Temporal Demand	The perceived time pressure or urgency experienced.
Performance	Self-assessment of task accomplishment or success
Effort	Subjective feeling of exertion or workload intensity
Frustration	Degree of annoyance, stress, or dissatisfaction experienced during the task.
RQ5: Group Dynamics	
Perceived communication efficiency	The level of resources consumed in performing tasks during the communication.
Perceived communication effectiveness	The communication among group members to complete tasks using the system and the perceived quality of those tasks.
Member activeness	The level of user motivation to engage in conversation.

There are various quantitative and qualitative techniques to measure the selected criteria. Post-experiment questionnaires will help answer the research questions RQ1, RQ2, RQ3, RQ4, and RQ5, designed according to the validation criteria. Questionnaires heavily rely on participants' opinions and the impossibility of ensuring that participants always provide objective answers. Their validity and reliability must be assessed carefully. Using questionnaires like SUS and NASA-TLX alleviates this problem since the scientific community validates them, and they have been accepted in many studies. It is crucial to take measurements to ensure that the survey maintains a decent level of validity and reliability. For the visual notation, we verified the validity and reliability of the questions since those questions were slightly altered. Furthermore, all questions are 5/7 points Likert scale in this thesis. We assume that the participants have similar perceptions of subjective scale responses.

4.2.1 Measure of User Satisfaction

User satisfaction includes usability, perceived efficiency, perceived effectiveness, perceived retention rate, intention to use in the future, usage fun, and user motivation.

Usability will be assessed through the most widely used standardised questionnaire, the System Usability Scale (SUS) (Brooke 1996). SUS is a Likert scale giving a global view of subjective usability assessments with ten items. We will use the 5-point scale from Strongly Disagree to Strongly Agree. The participants should be asked to mark their immediate response to each item without overthinking.

To get the final score of SUS, we would need to sum the score contributions from each item. The score contributions will range from 0 to 4 since we use the 5-point scale. The calculation of score contribution for each item, however, is different. If we mark the 5 points scale (Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree) with a scale position of (1,2,3,4,5) respectively. For items 1,3,5,7, and 9, the score contribution is the scale position minus 1; for items 2,4,6,8 and 10, the score contribution is five minus the scale position. Finally, we need to multiply the sum of the score contributions by 2.5 to obtain the overall value of the SUS score. Bangor et al. (Bangor, Kortum, and Miller 2008) proposed a grading scale that allows us to interpret the usability based on the SUS score. This grading scale assesses the sus score we obtained from the experimentation.

The 5-point Likert scale also assesses the rest of this category's criteria. Each criterion has one related question. The score of each criterion will be compared between the 2D modelling interface and the 3D modelling interface.

4.2.2 Measure of Visual Notation

Principles of Physics of Notation (Moody 2009) is used to assess the visual notation. The questions to evaluate PoN principles are adapted from the paper that examines the relationship between user perception of the quality of a conceptual modelling language from a cognitive point of view and its perceived usefulness (Figl and Derntl 2011). Each selected principle will have paired questions to verify the consistency of the answers. The questions are 5-point Likert Scale with adaption options for each question. (see Appendix E Visual Notation Questions) The principal score will be the mean of the two items. Comparison between the 2D modelling interface and the 3D immersive interface is based on the scores of each principle.

4.2.3 Measure of Presence

The Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, and Regenbrecht 2001) measures the sense of presence in the virtual environment. IPQ has three subscales: Spatial Presence, Involvement, Experienced Realism, and one additional general item that assesses the general "sense of being there". IPQ is a 7-point Likert scale with scale anchors like fully disagree to fully agree and some other scale anchors tailored according to the statements. Like different 7-point Likert scales, all items in IPQ range from 0 to 6 (0 on the left side and six

on the right side) except for Spatial Presence item 2, Involvement item 3, and Experienced Realism item 1. The range of 0 to 6 will be reversed for those three items. The score of each subscale will be the mean of each statement in that subscale.

4.2.4 Measure of Cognitive Load

NASA-TLX questionnaire is used to evaluate cognitive load. The study (Hart 2006) showed mixed results when comparing raw TLX scores to weighted TLX scores (some showed better sensitivity when removing weights, others showed no difference, and others showed less sensitivity). Since we have 15 mins for the questionnaire, we use the Raw NASA-TLX to eliminate the weighting process. The ratings are averaged based on the score contribution from 6 factors.

4.2.5 Measure of Group Dynamics

Two types of evaluation are conducted to assess group dynamics. The first type will be the data collected during the experiment, such as the total communicating time among the group and the number of times they change the fruits of their consensual labour. For the second type, similar to the perceived efficiency and perceived effectiveness in the category of user satisfaction, we proposed several questions to evaluate the perceived group dynamics. Those questions are presented as 5-point Likert scale questions (see Appendix I).

4.3 Online Survey

The validation process started with an online survey to evaluate the graphical modelling notation.

4.3.1 Research Questions

The first objective of this online survey is to get an initial idea of the visual differences between our 3D immersive interface and the classic diagrammatic interface using the principles of PoN (Moody 2009).

However, the 3D immersive interface involves factors other than the principles of PoN. One factor is the advantages of the 3D immersive MBSA interface for the spatial relations between graphical objects compared to the 2D diagrammatic MBSA interface. This factor may change the performance of MBSE activities.

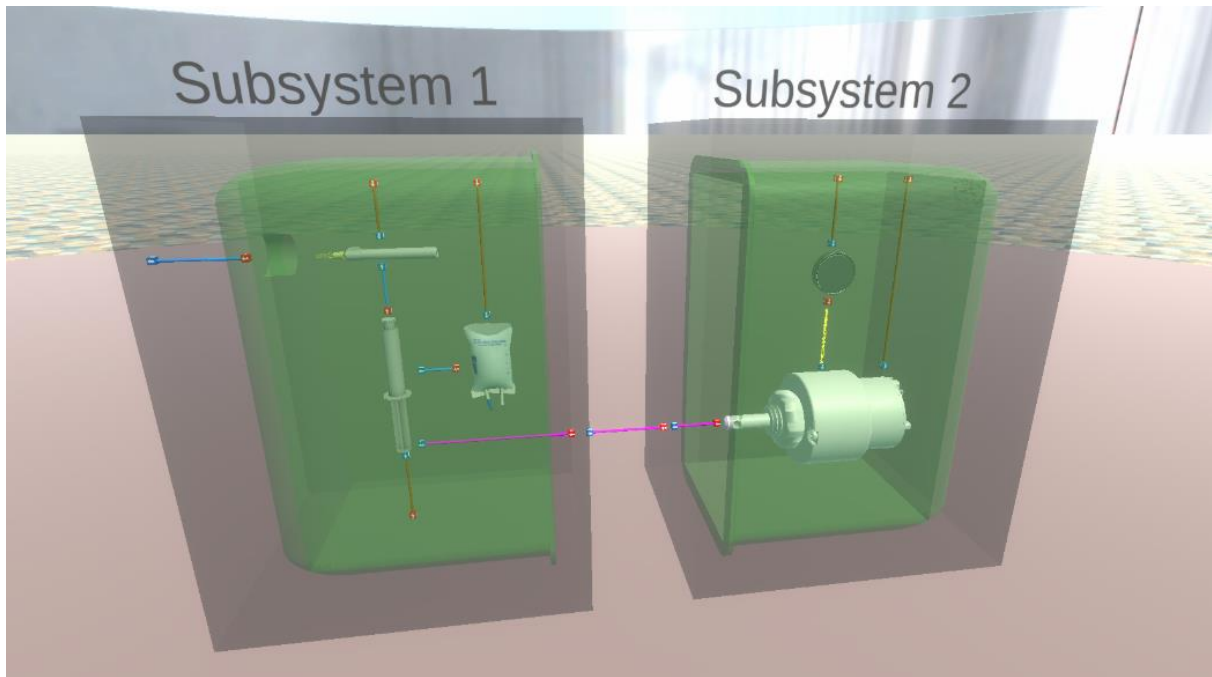


Figure 110 Illustration of spatial relation

Furthermore, the online survey aims to explore the impact of domain expertise level on the evaluation of visual notation. In the earlier section, we discussed one challenge concerning the difficulty of communication between the notation experts and non-experts. The cause behind this issue primarily stems from the nonexperts' unfamiliarity with abstract concepts in modelling languages, which is not a problem for the experts. Since this thesis aims to address this identified issue by proposing a 3D immersive MBSA interface, it becomes necessary to examine whether there is a difference in interpretation of the interface between experts and non-experts in the related field of the 3D immersive MBSA interface. This assessment is crucial for refining the research direction or improving the interface from diverse perspectives.

Therefore, the online survey includes the following research questions mentioned in section 1.3 RQ2+RQ2.1+RQ2.2:

Table 10 Validation methods and research questions (Survey)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey	RQ2+ RQ2.1+RQ2.2				
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

4.3.2 Online Survey Design

We designed the online survey¹ with a response time of approximately 15 minutes to avoid excessively long completion durations, as too-long surveys may result in significantly decreased response rates (Ozok 2012).

To answer research question 2.1, demographic questions are included in the survey at the beginning. Regarding the proposed 3D immersive interface, we have four demographic questions adapted from the OMG certificationⁱ about expertise level in MBSE, System Architecture, System Engineering, and Virtual Reality. (see Appendix D Demographic Questions)

What is your level of experience in Systems Architecting (in this survey, Systems architecting concentrates on the definition of the system lifecycle/context/stakeholders/interfaces/functions/requirements/... and the recursive process to decompose a system into sub-systems)?

- Novice:** I have never been involved in the definition of a system architecture.
- Awareness:** I have been involved in the definition of a system architecture on a toy problem.
- Supervised practitioner:** I have been involved in the definition of system architectures on industrial problems, but i require guidance and supervision.
- Practitioner:** I have participated in the architecture definition of several industrial systems. I am capable of guiding and advising others.
- Expert:** I have an extensive and substantial practical experience in systems architecting. I have defined or tailored processes, methods and tools for architecture definition within organizations. I can teach system architecting.

Next

Figure 111 Capture of the online survey demographic question

The next section of the questionnaire will be the questions about evaluating visual notation. Criteria are from Moody's PoN. (Moody 2009).

Seven representations will be evaluated using the PoN criteria. Four representations are SysML diagrams, including three symbolics and one iconic, where images aim to improve semantic transparency. Three representations of immersive 3D environments are captured.

¹ <https://www.questionpro.com/a/TakeSurvey?tt=cQzaK3hY%2BvgEChrPeIW9eQ%3D%3D>

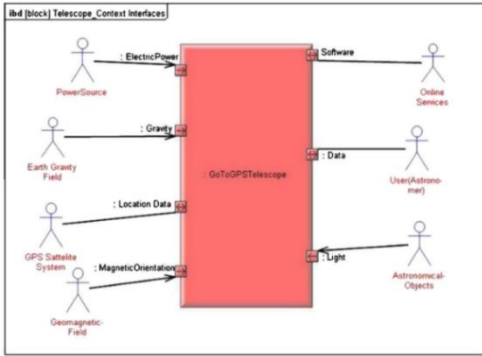


Figure 112 Representation 1

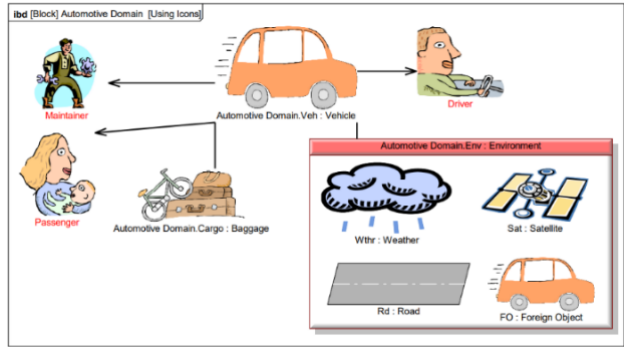


Figure 113 Representation 2

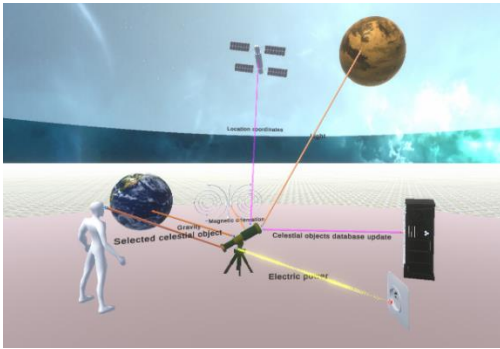


Figure 114 Representation 3

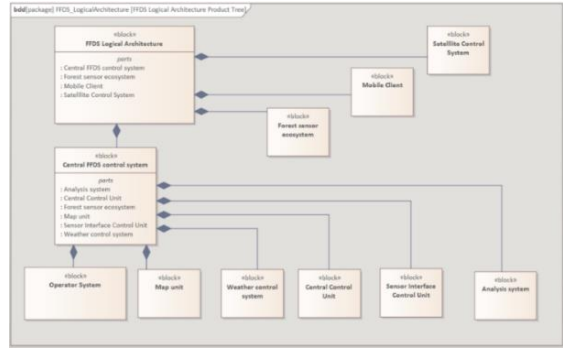


Figure 115 Representation 4

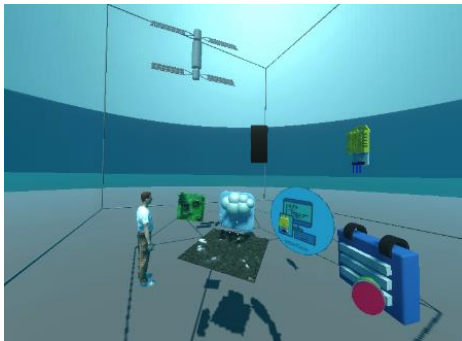


Figure 116 Representation 5

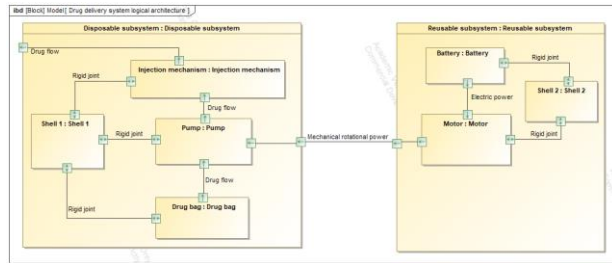


Figure 117 Representation 6

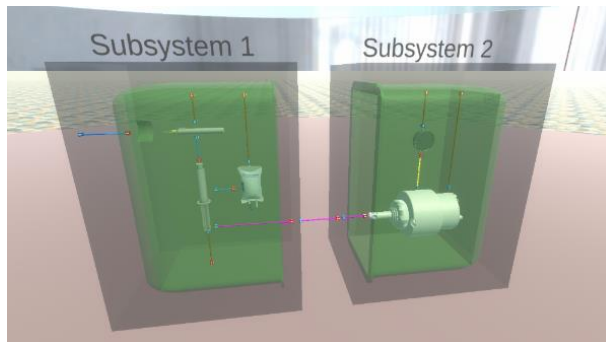


Figure 118 Representation 7

In total, there are 30 questions to evaluate the visual notation. Each criterion will be evaluated three times since there are two questions for each criterion. (5 criteria * 2 paired questions * 3 times = 30 questions, see Table 11)

Table 11 Representation and PoN Criteria Mapping

	Perceptual Discriminability	Semantic Transparency	Complexity Management	Visual Expressiveness	Graphic Economy
Rep1 (2D symbolic)	X				X
Rep2 (2D Iconic)			X	X	
Rep3 (3D)		X	X	X	
Rep4 (2D symbolic)		X	X		
Rep5 (3D)	X				X
Rep6 (2D symbolic)		X		X	
Rep7 (3D)	X				X

This way, each criterion is evaluated three times, either with one 2D interface and two 3D interfaces or two 2D interfaces and one 3D interface.

Additionally, with the last two representations of the drug delivery system case study, one extra question related to research question 2.2 about the spatial relation.

4.3.3 Online Survey Validation

During the design phase of a survey, it is crucial to verify its validity and reliability. Validity is the degree to which the survey instrument is valid in the results it produces or, in other words, whether the survey is measuring what it says it is measuring (accuracy). Reliability is the extent to which a measurement of a phenomenon provides stable and consistent results (consistency). Figure 119 depicts those two concepts. Scores on the top left are both reliable and valid. Scores on the top right are highly reliable but not valid, demonstrating systematic error. Scores on the bottom left are highly valid but need to be more reliable. Scores on the bottom right are neither reliable nor valid, demonstrating random error (Lundberg 2006).

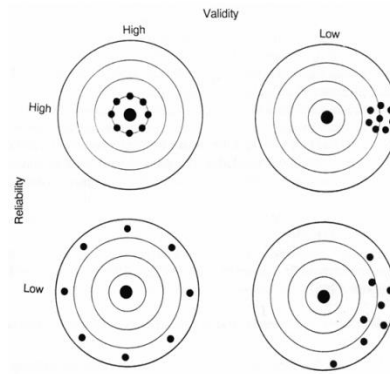


Figure 119 The relationship between reliability and validity is illustrated as “target diagrams” (Lundberg 2006)

Validity:

Validity is categorised into four types: face validity, content validity, construct validity, and criterion validity (see Figure 120).

- “Face validity is a subjective judgment on the operationalisation of a construct. It is the degree to which a measure appears to be related to a specific construct, in the judgment of non-experts such as test takers and legal system representatives” (Taherdoost 2016).
- Content validity is “the degree to which items in an instrument reflect the content universe to which the instrument will be generalised” (Straub, Boudreau, and Gefen 2004).
- “Construct validity refers to how well you translated or transformed a concept, idea, or behaviour that is a construct into a functioning and operating reality, the operationalisation” (Taherdoost 2016).
- “Criterion validity is the extent to which a measure is related to an outcome. It measures how well one measure predicts an outcome for another measure” (Taherdoost 2016).

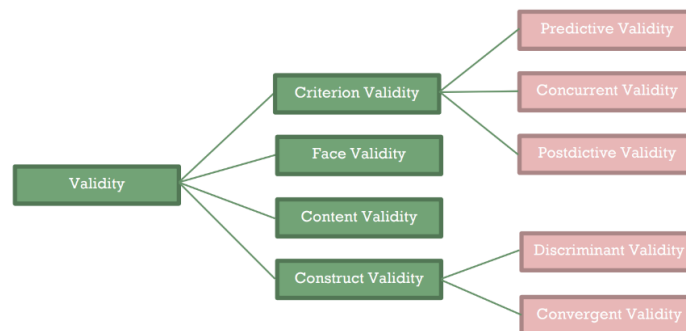


Figure 120 Subtypes of various forms of validity tests (Taherdoost 2016)

Reliability:

Reliability refers to the extent to which a measurement provides stable and consistent results (Carmines and Zeller, 1979). It is also about repeatability as well. “A scale or test is said to be reliable if repeat measurement made by it under constant conditions will give the same result” (Moser and Kalton 2017).

For the survey conducted in this thesis, we don't aim to provide a survey with very scientifically grounded results but a first peer of the performance of the 3D immersive MBSA interface. A significant part of the items in the questionnaire is adapted from (Figl and Derntl 2011), except for the criterion Complexity Management. They argued that participants need to know more about diagram types and their relationships to evaluate cognitive fit, complexity management, and cognitive integration. However, according to Moody's paper (Moody 2009) proposing the PoN, complexity management is the capacity of a visual notation to represent information without overloading the human mind. There are mechanisms to help improve Complexity Management, such as Modularization and Hierarchy (Moody 2009). The complexity management can be reflected in the diagram(s) by evaluating the mental load related to the diagram(s). Therefore, this criterion is included in this survey.

A pilot survey was conducted to ensure content validity. Researchers reviewed each question to identify confusing or misleading questions and validate that the survey measures what it claims to measure. The scale items were refined based on the question so that participants could intuitively select items. For instance, for the question: “*The graphical objects are*”, the scale items “*Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly Disagree*” are refined into “*Extremely easy to differentiate; Easy to differentiate; Neither easy nor difficult to differentiate; Difficult to differentiate; Extremely easy to differentiate*”.

As we used the questions from (Figl and Derntl 2011), a Cronbach's alpha test was used to test the reliability, as in the study (Figl and Derntl 2011). Each criterion has dual questions. Cronbach's alpha is used to check if the dual questions are reliable, i.e., adequate internal consistency for the scales. However, due to the context limitation and time constraints, it wasn't easy to find 10 participants in advance for the Cronbach's alpha test to ensure reliability. The test was conducted after the diffusion, and the results were used to proceed the Cronbach's alpha test. Table 12 shows the results for each criterion.

Table 12 Results of Cronbach's Alpha for each PoN criteria

	Cronbach's Alpha
Perceptual Discriminability	.645
Semantic Transparency	.719
Complexity Management	.340
Visual Expressiveness	.731
Graphic Economy	.580

The standard threshold for alpha is $\alpha > .6$ or $.7$, but the length of scale can influence the value of alpha. Longer scale lengths, such as those with many scale items, give higher alpha values. A small number of scale items would provide a lower reliability coefficient, which is the case for some criteria in this survey, for instance, five metric scales (*“Extremely easy to differentiate; Easy to differentiate; Neither easy nor difficult to differentiate; Difficult to differentiate; Extremely easy to differentiate”*) or three metric scale items in the symmetric situation (*“Significantly reducing the variation of colour, size, form or brightness; Reducing variation of color, size, form or brightness; No need for improvement; increasing variation of color, size, form or brightness; Significantly increasing variation of color, size, form or brightness”*). Moreover, studies (Hinton, McMurray, and Brownlow 2014) indicate that an alpha from $.5$ to $.7$ could show moderate reliability. Thus, with Complexity management giving an alpha value lower than $.5$, we use only one item to evaluate it. Moreover, we calculated the results in the later analysis using both objects and one single item of Complexity Management. Even the outcome of using a single item differs from using both things (the difference between them crosses the set-up significance level of $.005$). However, the result of the p-value was still very close, with an alpha value of $.340$. This observation corroborates the acceptable reliability of low alpha with fewer scale items.

4.3.4 Online Survey Diffusion

The survey is diffused by e-mails to the INCOSE French chapter AFIS, the Design Society, the Discord Systems Engineering server, which gathers an international community of systems engineers, and various systems engineering groups on LinkedIn (OMG group, INCOSE group, AFIS group, etc.). The survey targets industrial practitioners familiar with system engineering and students in the first year of a bachelor's degree in industrial engineering who will mostly be novices to systems engineering and virtual reality. The context constraints make setting up a sample size and diffusing the survey impossible, as finding enough experts and students is challenging. The strategy was to publicly diffuse the online survey and collect as much data as possible by the end of data collecting.

4.3.5 Online Survey Analysis

Survey analysis is presented first before the results are presented, leading to less biased results. Survey analysis aims to exclude invalid data and decide on statistical tests that address the research questions. To begin with the survey analysis, the first step is to exclude invalid responses. A response is weak when the box of the consent form is not checked, or the participants answered fewer than four questions. A threshold of four questions was chosen because, according to the responses, we found six responses with less than four, then at least eight. We received 118 responses after excluding invalid responses.

The next step is to remove outliers to avoid their negative influence on the results. Since the questionnaire is based on an exploratory study, we could not make clear expectations about the results. The metrics the questionnaire evaluates are based on the participants' perceptions. This echoes the important assumption throughout the survey: we assume that participants can correctly estimate metric values based on the perceived information, implying that perceived metric values approximate actual metric values. This means that even if we notice some outlier values with the boxplot, we can't exclude those outliers since it could be just how the participant perceived the given representation—assuming, of course, that we have ruled out the possibility of data errors (transcription errors or recode errors). However, we can govern the “Mischievous participants” type outliers by the time taken to complete the questionnaire. The average time taken is 780s (13 mins). The extremely long time taken is reasonable since the participant could be interrupted by something else during the session and return to the survey in a while without closing it. However, it is not possible that a participant took 170 seconds to answer all 36 questions, plus the introduction part to read. We set up a time limit of 300 seconds for the complete response, which means if the participant used less than 300 seconds to answer all 36 questions, he probably randomly selected the answers. In this case, the response is considered invalid. Two responses were removed after this step (205s for 36 questions and 170s for 36 questions). Thus, after filtering outliers, we have responses in total 116.

As we evaluate each principle of PoN three times on different representations, for research question 2, we will conduct a one-way repeated measures ANOVA to check for a significant difference in each criterion. If there is, we will proceed with a post-hoc test with Bonferroni to adjust pairwise comparisons. We have an interface as an independent variable for a one-way repeated measure ANOVA test, and each validation criterion will be a dependent variable.

As for the independent variable, we have a 2D diagrammatic MBSA interface, a 2D iconic MBSA interface, or a 3D immersive MBSA interface (2D captures of 3D immersive MBSA interface). For some principles, we may only be able to compare a 2D diagrammatic MBSA interface and a 3D immersive MBSA interface, with either two 2D diagrammatic interfaces or

two 3D immersive interfaces with different case studies. For instance, perceptual discriminability is evaluated on one 2D symbolic (Rep1) and two 3D (Rep5 and Rep7) (see Table 13).

Table 13 Representation and PoN Criteria Mapping

	Perceptual Discriminability	Semantic Transparency	Complexity Management	Visual Expressiveness	Graphic Economy
Rep1 (2D symbolic)	X				X
Rep2 (2D Iconic)			X	X	
Rep3 (3D)		X	X	X	
Rep4 (2D symbolic)		X	X		
Rep5 (3D)	X				X
Rep6 (2D symbolic)		X		X	
Rep7 (3D)	X				X

The test planning will be slightly twisted for research question 2.1 (impact of expertise level on evaluation). We want an overview of how the expertise level will influence the criteria assessment. The first step will be to observe if the expertise level in related domains will affect the criteria evaluation on the 2D diagrammatic interface and 3D immersive interface. For each criterion, there are three representations with different interfaces. When there are two 3D immersive interfaces or two 2D diagrammatic interfaces, we will calculate their means to conduct an independent sample t-test with the new mean value only once instead of doing it twice with each of them. Since there are four expertise levels in related domains, there will be four evaluations (with each expertise level) on the 5 PoN criteria. Due to the limitation of the sample size, we recorded the response of expertise level into two levels: non-expert (“*Novice; User; Awareness*”) and expert (“*Builder; Practitioner; Expert*”).

For research question 2.2 about the spatial relation factor, as there are only two groups (2D symbolic vs. 3D), we will conduct a paired sample t-test to determine if there is a significant difference. Theoretically, a 3D immersive interface aims to bring more information about spatial relations. This means the one-tailed test will be the favourable test here. However, it doesn’t automatically eliminate the possibility that the 3D immersive interface may provide worse results of spatial relations with different interpretations of spatial relations from users. If this is the case, we will not be able to detect this outcome with a one-tailed test since a one-tailed test won’t distinguish between “worse” and “not significantly different”. Therefore, to avoid p-hacking, using the two-tailed test is better. Suppose the p-value of the two-tailed test is more significant than 0.05. In that case, we fail to reject the null hypothesis, meaning no significant difference exists between the 3D immersive and 2D diagrammatic interfaces. Suppose the p-

value of the two-tailed test is smaller than 0.05, which means there is a significant difference between the 3D immersive interface and the 2D diagrammatic interface. In that case, we can reject the null hypothesis, and then, based on the sign of the t-value, we would know the direction, that is, which interface provides the correct spatial relation.

4.3.6 Online Survey Results

This section reports the results of the online survey: one-way repeated measures ANOVA test for research question 2: Does the 3D immersive MBSA interface provide better visual notations regarding the Physics of Notation (PoN) than the 2D diagrammatic MBSA interface; independent sample t-test for research question 2.1: Does the expertise level in MBSA/SE/SA/VR influence the evaluation of visual notation; paired sample t-test for research question 2.2: Does the extra information about the spatial relation provided by the 3D immersive interface improve the MBSE activities.

Paired sample t-test for research question 2:

RQ2: Does the 3D immersive MBSA interface provide better **visual notations** regarding the Physics of Notation (PoN) than the 2D diagrammatic MBSA interface?

- **Perceptual Discriminability:**

The independent variable is the type of modelling interface, which has three modalities: a 2D diagrammatic interface (Telescope), a 3D immersive interface (Forest Fire Detection System), and a 3D immersive interface (Drug Delivery System).

Table 14 Within-Subjects Factors of ANOVA test for Perceptual Discriminability

Measure: Perceptual Discriminability	
Interface	Dependent Variable
1	PD_2DTelescope
2	PD_3DFFDS
3	PD_3DDDS

Table 15 Descriptive Statics of ANOVA test for Perceptual Discriminability

	Mean	Std. Deviation	N
PD_2DTelescope	3,0849	,83800	106
PD_3DFFDS	2,8962	,91734	106
PD_3DDDS	3,3208	,83435	106

The results of ANOVA indicated a significant difference, Wilks' Lambda=.871, $F(2,104) = 7.728$, $p < .001$, $\eta^2 = .129$, which means we can reject the null hypothesis that there are no significant differences in Perceptual discriminability between the different interface types.

Table 16 Multivariate Tests of ANOVA test for Perceptual Discriminability

	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Interface	Pillai's Trace	,129	7,728b	2,000	104,000	<,001	,129
	Wilks' Lambda	,871	7,728b	2,000	104,000	<,001	,129
	Hotelling's Trace	,149	7,728b	2,000	104,000	<,001	,129
	Roy's Largest Root	,149	7,728b	2,000	104,000	<,001	,129

a. Design: Intercept Within Subjects Design: Interface

b. Exact statistic

Following up with the pairwise comparison between each group, we found out that there is no significant difference between 2D diagrammatic interface and 3D immersive interface (FFDS), nor between 2D diagrammatic interface and 3D immersive interface (DDS). The critical difference comes from two 3D immersive interfaces with $p < .001$.

Table 17 Pairwise Comparisons of ANOVA test for Perceptual Discriminability

Measure: Perceptual Discriminability

(I) Interface	(J) Interface	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	,189	,127	,425	-,121	,499
	3	-,236	,111	,107	-,505	,034
2	1	-,189	,127	,425	-,499	,121
	3	-,425*	,110	<,001	-,693	-,156
3	1	,236	,111	,107	-,034	,505
	2	,425*	,110	<,001	,156	,693

Based on estimated marginal means

*. The mean difference is significant at the 05 level.

b. Adjustment for multiple comparisons: Bonferroni.

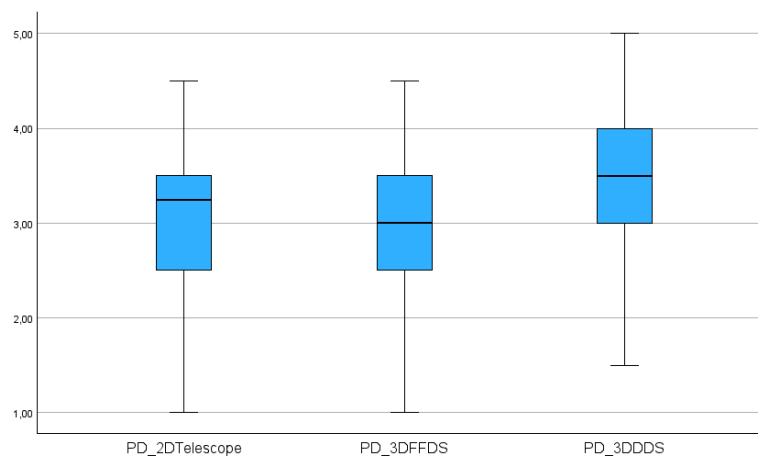


Figure 121 Boxplot of the Perceptual Discriminability result

The result indicates that the Perceptual Discriminability of the 3D immersive interface can fluctuate considerably depending on the system of interest. In this case, the reasons for the Perceptual Discriminability variation may come from diverse sources. For example, the number of elements or the level of abstraction of the components in two different case studies may cause the difference. Similarly, those are the reasons that we could have provided better perceptual discriminability with a 3D immersive interface than the 2D diagrammatic interface.

- **Semantic Transparency:**

The independent variable is the type of modelling interface with two modalities: 3D immersive interface (Telescope), 2D diagrammatic interface (Forest Fire Detect System), and 2D diagrammatic interface (Drug Delivery System).

Table 18 Within-Subjects Factors of ANOVA test for Semantic Transparency

Measure: Semantic Transparency

Interface	Dependent Variable
1	ST_3DTelescope
2	ST_2DFFDS
3	ST_2DDDS

Table 19 Descriptive Statistics of ANOVA test for Semantic Transparency

	Mean	Std. Deviation	N
ST_3DTelescope	3,4439	,88575	107
ST_2DFFDS	2,9112	,99483	107
ST_2DDDS	3,1075	,99534	107

The result of ANOVA indicated a significant difference, Wilks' Lambda=.870, $F(2.105)=7.864$, $p<.001$, $\eta^2=.130$. Thus, there is significant evidence to reject the null hypothesis.

Table 20 Multivariate Tests of ANOVA test for Semantic Transparency

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Interface Pillai's Trace	,130	7,864 ^b	2,000	105,000	<,001	,130
Wilks' Lambda	,870	7,864 ^b	2,000	105,000	<,001	,130
Hotelling's Trace	,150	7,864 ^b	2,000	105,000	<,001	,130
Roy's Largest Root	,150	7,864 ^b	2,000	105,000	<,001	,130

a. Design: Intercept Within Subjects Design: Interface

b. Exact statistic

Follow-up comparisons indicated a significant difference between the 3D immersive interface (Telescope) and 2D diagrammatic interface (FFDS) with $p<.001$, a significant difference between the 3D immersive interface (Telescope) and 2D diagrammatic interface

(DDS) with $p=.045$, no significant difference between 2D diagrammatic interface (FFDS) and 2D diagrammatic interface (DDS) with $p=.086$.

Table 21 Pairwise Comparisons of ANOVA test for Semantic Transparency

Measure: Semantic Transparency

(I) Interface	(J) Interface	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	,533*	,138	<,001	,198	,868
	3	,336*	,136	,045	,006	,667
2	1	-,533*	,138	<,001	-,868	-,198
	3	-,196	,089	,086	-,412	,019
3	1	-,336*	,136	,045	-,667	-,006
	2	,196	,089	,086	-,019	,412

Based on estimated marginal means

*. The mean difference is significant at the,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

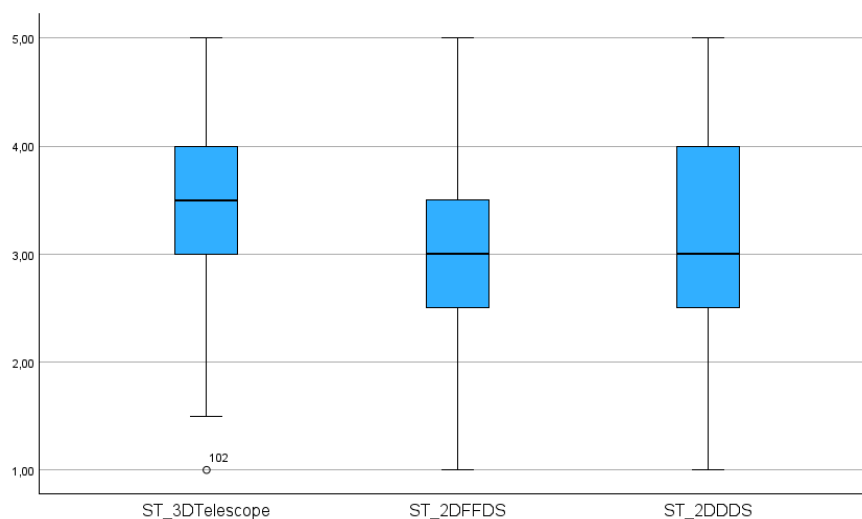


Figure 122 Boxplot of the Semantic Transparency result

- **Complexity Management:**

The independent variable is the type of modelling interface, which has three modalities: 2D iconic interface (Automotive), 3D immersive interface (Telescope), and 2D diagrammatic interface (Forest Fire Detection System).

Table 22 Within-Subjects Factors of ANOVA test for Complexity Management

Measure: Complexity Management

Interface	Dependent Variable
1	CM_2DAutomotive
2	CM_3DTelescope

Table 23 Descriptive Statistics of ANOVA test for Complexity Management

	Mean	Std. Deviation	N
CM_2DAutomotive	3,9630	1,06715	108
CM_3DTelescope	3,7778	,96995	108
CM_2DFFDS	3,4167	,91840	108

The result of ANOVA indicated a significant difference, Wilks' Lambda=.866, $F(2.106)=8.231$, $p<.001$, $\eta^2=.134$. There is significant evidence to reject the null hypothesis.

Table 24 Multivariate Tests of ANOVA test for Complexity Management

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	
Interface	Pillai's Trace	,134	8,231 ^b	2,000	106,000	<,001	,134
	Wilks' Lambda	,866	8,231 ^b	2,000	106,000	<,001	,134
	Hotelling's Trace	,155	8,231 ^b	2,000	106,000	<,001	,134
	Roy's Largest Root	,155	8,231 ^b	2,000	106,000	<,001	,134

a. Design: Intercept Within Subjects Design: Interface

b. Exact statistic

Follow-up comparisons indicated no significant difference between the 2D diagrammatic iconic interface (Automotive) and 3D immersive interface (Telescope) with $p=.390$, a significant difference between the 3D immersive interface (Telescope) and 2D diagrammatic symbolic interface (FFDS) with $p=.020$, a significant difference between 2D diagrammatic iconic interface (Automotive) and 2D diagrammatic symbolic interface (FFDS) with $p<.001$.

Table 25 Pairwise Comparisons of ANOVA test for Complexity Management

Measure: Complexity Management

(I) Interface	(J) Interface	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	,185	,121	,390	-,110	,480
	3	,546*	,135	<,001	,218	,875
2	1	-,185	,121	,390	-,480	,110
	3	,361*	,131	,020	,044	,679
3	1	-,546*	,135	<,001	-,875	-,218
	2	-,361*	,131	,020	-,679	-,044

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

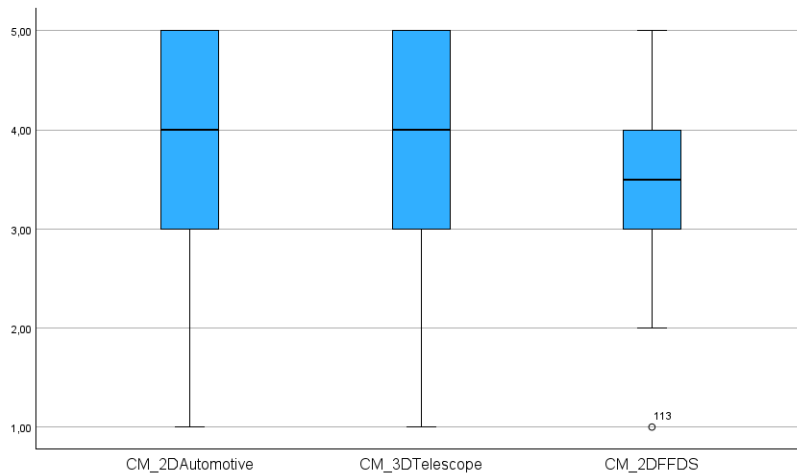


Figure 123 Boxplot of the Complexity Management result

- **Visual Expressiveness:**

The independent variable is the type of modelling interface with three modalities: 2D iconic interface (Automotive), 3D immersive interface (Telescope), and 2D diagrammatic interface (Drug Delivery System).

Table 26 Within-Subjects Factors of ANOVA test for Visual Expressiveness

Interface	Measure: Visual Expressiveness Dependent Variable
1	VE_2DAutomotive
2	VE_3DTelescope
3	VE_2DDDS

Table 27 Descriptive Statistics of ANOVA test for Visual Expressiveness

	Mean	Std. Deviation	N
VE_2DAutomotive	3,5802	,88945	106
VE_3DTelescope	3,7170	,93100	106
VE_2DDDS	3,1887	1,06336	106

The result of ANOVA indicated a significant difference, Wilks' Lambda=.900, $F(2,104)=5.768$, $p=0.04$, $\eta^2=.100$. There is significant evidence to reject the null hypothesis.

Table 28 Multivariate Tests^a of ANOVA test for Visual Expressiveness

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Interface Pillai's Trace	,100	5,768b	2,000	104,000	,004	,100
Wilks' Lambda	,900	5,768b	2,000	104,000	,004	,100
Hotelling's Trace	,111	5,768b	2,000	104,000	,004	,100
Roy's Largest Root	,111	5,768b	2,000	104,000	,004	,100

a. Design: Intercept Within Subjects Design: Interface

b. Exact statistic

Follow-up comparisons indicated no significant difference between the 2D iconic interface (Automotive) and 3D immersive interface (Telescope) with $p=.603$, a significant difference between the 3D immersive interface (Telescope) and 2D diagrammatic interface (DDS) with $p=.003$, a substantial difference between the 2D iconic interface (Automotive) and 2D diagrammatic interface (DDS) with $p=.020$.

Table 29 Pairwise Comparisons of ANOVA test for Visual Expressiveness

Measure: Visual Expressiveness

(I) Interface	(J) Interface	Mean Difference (I-J)	Std. Error	Sig. b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	-,137	,106	,603	-,395	,122
	3	,392*	,142	,020	,047	,736
2	1	,137	,106	,603	-,122	,395
	3	,528*	,156	,003	,150	,907
3	1	-,392*	,142	,020	-,736	-,047
	2	-,528*	,156	,003	-,907	-,150

Based on estimated marginal means

*. The mean difference is significant at the 05 level.

b. Adjustment for multiple comparisons: Bonferroni.

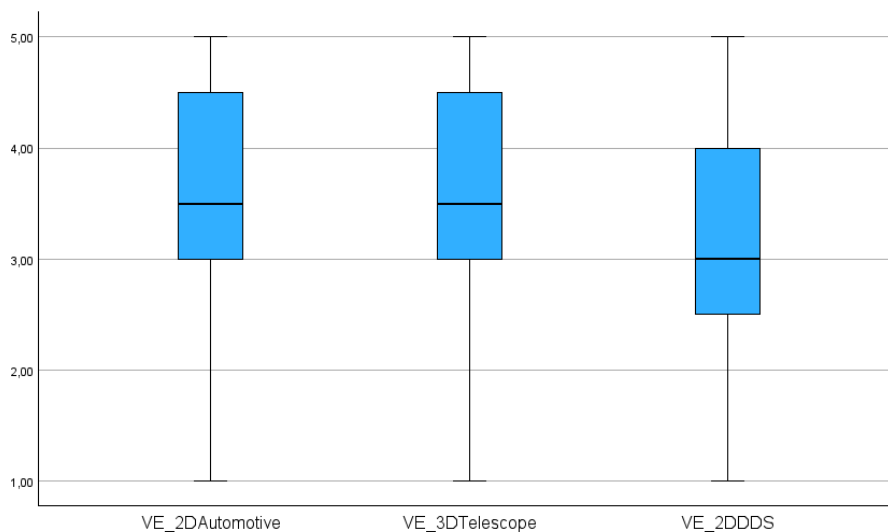


Figure 124 Boxplot of the Visual Expressiveness result

- **Graphic Economy**

The independent variable is the type of modelling interface, which has two modalities: a 2D diagrammatic interface (Telescope), a 3D immersive interface (Forest Fire Detection System), and a 3D diagrammatic interface (Drug Delivery System).

Table 30 Within-Subjects Factors of ANOVA test for Graphic Economy

Measure: Graphic Economy

Interface Dependent Variable

1	GE_2DTelescope
2	GE_3DFFDS
3	GE_3DDDS

Table 31 Descriptive Statistics of ANOVA test for Graphic Economy

	Mean	Std. Deviation	N
GE_2DTelescope	3,3143	1,06802	105
GE_3DFFDS	3,2857	1,11557	105
GE_3DDDS	4,0857	,93144	105

The result of ANOVA indicated a significant difference, Wilks' Lambda=.602, $F(2.103)=34.014$, $p<.001$, $\eta^2=.398$. There is significant evidence to reject the null hypothesis.

Table 32 Multivariate Tests^a of ANOVA test for Graphic Economy

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Interface Pillai's Trace	,398	34,014 ^b	2,000	103,000	<,001	,398
Wilks' Lambda	,602	34,014 ^b	2,000	103,000	<,001	,398
Hotelling's Trace	,660	34,014 ^b	2,000	103,000	<,001	,398
Roy's Largest Root	,660	34,014 ^b	2,000	103,000	<,001	,398

a. Design: Intercept

Within Subjects Design: Interface

b. Exact statistic

Follow-up comparisons indicated no significant difference between the 2D diagrammatic interface (Telescope) and 3D immersive interface (FFDS) with $p=1.000$, a substantial difference between 3D immersive interface (FFDS) and 3D immersive interface (DDS) with $p<.001$, a significant difference between 2D diagrammatic interface (Telescope) and 3D immersive interface (DDS) with $p<.001$.

Table 33 Pairwise Comparisons of ANOVA test for Graphic Economy

Measure: Graphic Economy

(I) Interface	(J) Interface	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	,029	,166	1,000	-,375	,433
	3	-,771*	,137	<,001	-1,104	-,438
2	1	-,029	,166	1,000	-,433	,375
	3	-,800*	,117	<,001	-1,084	-,516
3	1	,771*	,137	<,001	,438	1,104
	2	,800*	,117	<,001	,516	1,084

Based on estimated marginal means

*. The mean difference is significant at the 05 level.

b. Adjustment for multiple comparisons: Bonferroni.

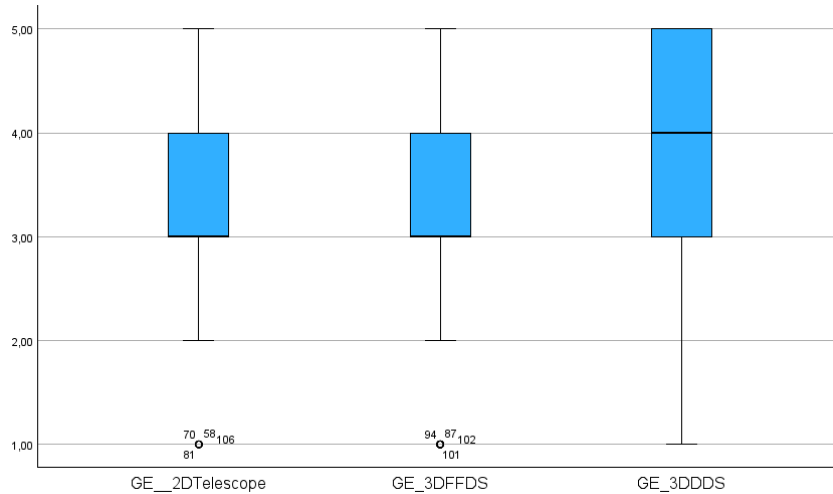


Figure 125 Figure 15 Boxplot of the Graphic Economy result

Like perceptual discriminability, we have a similar result here with graphic economy. We could have improved this quality criterion with the Forest Fire Detect System compared to the 2D diagrammatic interface, but it was significantly enhanced with the Drug Delivery System.

Independent sample t-test for research question 2.1:

RQ 2.1: Does the expertise level in MBSA/SE/SA/VR influence the evaluation of visual notation?

We run an independent sample t-test for each criterion with different expertise levels in the related domains. Most groups show no significant difference, with very few groups showing a significant difference. Here is a sheet summarising all the independent sample t-tests:

Table 34 Results of independent sample t-test

MBSE	SA	Perceptual Discriminability		Semantic Transparency		Complexity Management	
		SE	VR				
2D Diagrammatic		t=2.111 p=.037*	t=1.993 p=.049*	t=-3.257 p=.002*	t=-1.211 p=.229	t=-.635 p=.527	t=-.356 p=.723
		t=2.537 p=.013*	t=4.493 p<.001*	t=.068 p=.946	t=-.393 p=.695	t=-1.527 p=.130	t=.118 p=.906
2D Iconic						t=.276 p=.783	t=-1.139 p=.257
						t=.281 p=.779	t=1.430 p=.155
3D Immersive		t=.589 p=.557	t=1.950 p=.054	t=-1.388 p=.168	t=.124 p=.901	t=-.044 p=.965	t=.325 p=.746
		t=.983 p=.328	t=-.548 p=.585	t=-.374 p=.709	t=-.422 p=.674	t=-.516 p=.607	t=.078 p=.938
MBSE	SA	Visual Expressiveness			Graphic Economy		
SE	VR						

2D Diagrammatic	t=-.716 p=.476	t=.160 p=.873	t=-.496 p=.621	t=-1.826 p=.070
	t=.807 p=.421	t=1.247 p=.215	t=-1.359 p=.177	t=.937 p=.351
2D Iconic	t=-.106 p=.988	t=-1.305 p=.195		
	t=.222 p=.825	t=1.596 p=.113		
3D Immersive	t=.017 p=.986	t=.286 p=.775	t=.779 p=.438	t=.202 p=.840
	t=-.295 p=.768	t=-.302 p=.763	t=-.377 p=.707	t=-.128 p=.899

Paired sample t-test for research question 2.2

Research question 2.2: Does the extra information about the **spatial relation** provided by the 3D immersive interface improve the MBSE activities?

The independent variable is the type of modelling interface, which has two modalities: a 2D diagrammatic interface and a 3D immersive interface.

Table 35 Paired Samples Statistics of paired sample t-test for spatial relation

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SR_2D	,8774	106	,83612	,08121
	SR_3D	1,5755	106	,70303	,06828

The result indicates a marginally significant difference in the spatial relation carried out by both interfaces, $t=-8.217$, $df=105$, $p<.001$.

Table 36 Paired Samples Test of paired sample t-test for spatial relation

	Paired Differences					Significance			
	Mean	Std. Deviation	Std. Error	95% Confidence Interval of the Difference		t	df	One-Sided p	Two-Sided p
				Lower	Upper				
SR_2D - SR_3D	-,69811	,87472	,08496	-,86657	-,52965	-8,217	105	<,001	<,001

The results in Figure 126 imply that the 3D immersive interface makes people more likely to interpret spatial relationships similarly. This could help minimise information misinterpretation in the early stages of systems design.

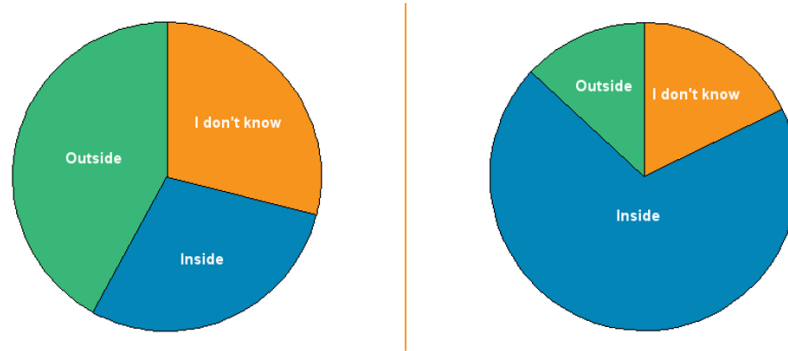


Figure 126 Pie chart on the percentage of the answers (2D left, 3D right)

4.3.7 Online Survey Discussion

In summary, based on what the survey has tested, the 3D immersive interface effectively improves Semantic Transparency, Complexity Management, and Visual Expressiveness with a significant difference from the 2D diagrammatic symbolic interface. This conclusion lines in with our expectations. For example, a more intuitive representation metaphor for the elements carried out by the interface using 3D models instead of always presenting them with boxes will vastly reduce the perception and interpretation barrier. This transformation from semantically opaque to semantically immediate contributes to a very intuitive qualitative increase in semantic transparency. A similar effect can be concluded for visual expressiveness since the 3D immersive interface possesses much more diversity (almost unlimited) of visual variables, naturally making it have better visual expressiveness. Such modifications on the visual graphics impact complexity management, where the presented information is easier to understand, leading to significantly better complexity management.

The 2D diagrammatic iconic interface has some improvements in a similar direction to the 3D immersive interface compared to the 2D diagrammatic symbolic interface. They provide almost equal complexity management. The 3D immersive interface has a slightly better visual expressiveness than the 2D diagrammatic iconic interface. This result indicates no extensive range of improvement for complexity management and visual expressiveness from a 2D diagrammatic iconic interface to a 3D immersive interface.

However, we could have improved the perceptual discriminability and graphic economy with the 3D immersive interface. This may be strongly related to the elements in the case study with different quantity levels and whether the objects to represent are concrete or abstract. For instance, finding a 3D chair model to represent a concrete (i.e., tangible) concept like a Chair is easy. However, when submitting an abstract concept that is physically invisible in the real world (e.g., emotions, thoughts, ideas), sometimes it would be impossible to find a 3D model

to represent it so everyone will recognise it. Therefore, it is more challenging to represent an abstract concept even with a 3D immersive interface (i.e., a 3D model for an abstract concept with good semantic transparency). Consequently, participants won't be able to readily receive the information the elements carry, let alone distinguish between different objects. On the other hand, a 3D immersive interface brings more data to facilitate the "diagrammatic reasoning" process. Still, in the meantime, it makes it much easier to overload the human mind, making the whole environment overwhelming. This brings out another observation: perceptual discriminability and graphic economy vary dramatically in 3D immersive interfaces when the system of interest changes. This implies that those metrics are highly responsive to the projects conducted in a 3D immersive interface. 3D immersive interfaces with different projects may come out with entirely different performances. 2D diagrammatic symbolic interface, on the other hand, is much more robust than 3D immersive interface. For the evaluation of semantic transparency, there are two different case studies (i.e., one case study of a Forest Fire Detection System and another case study of a Drug Delivery System) with 2D diagrammatic symbolic interface, and the results demonstrate nearly identical outcomes, with minimal variations in the means between those two case studies. Due to the time limit of the online survey, we could not evaluate every criterion in all different interfaces with different case studies, which means we cannot know if this phenomenon will be present in all assessment criteria across other interfaces. This can potentially be one part of future work, but for now, what we can already conclude is that the advantage of the 3D immersive interface is not always practical, depending on the case study. Based on this first validation step, we can argue that a 3D immersive interface provides a better global visual notation than the 2D diagrammatic symbolic interface. Still, the range of differences depends on the project content.

The spatial relation factor was proved well presented with a 3D immersive interface to have common interpretations from users, and it can play an important role when this spatial relation factor is more crucial for the interested system. However, this conclusion may be anecdotic since there was only one case study with two representations tested for the spatial relation. A detailed study covering more case studies focusing on spatial relations may be conducted in the future for a rounded analysis.

As for the impact of different expertise levels in related domains on evaluating the criteria of PoN, there are only a few significant differences in the results. The significant differences only appear in the 2D diagrammatic symbolic interface, including the MBSE/SA/SE/VR expertise level with Perceptual Discriminability and the MBSE expertise level with Semantic Transparency. Statistical tests show that experts in MBSE/SA/SE/VR are more likely to give a lower score on Perceptual Discriminability in 2D diagrammatic symbolic interface than novices, which is an astonishing result. One assumption for this outcome is that the experts used to

work with complicated diagrams more frequently than novices. Hence, to work more efficiently, they raise the bar of this metric, which means they bring their own experience in their area of specialisation and subconsciously incorporate their own “biased viewpoints” in the measurement process. This leads to one limitation of the online survey: the measured quality criteria rely on participants’ perceptions rather than an objective truth factor. Further research on how they evaluate it can help us better understand this strange phenomenon. Regarding Semantic Transparency, it’s understandable that experts tend to give a higher score for Semantic Transparency in 2D diagrammatic symbolic interfaces than novices. Indeed, they are way more familiar with the visual syntax behind the semantic constructs and know the exact meaning of the bewildering array of boxes and arrows of SysML. The significant difference here also highlights the interest in 3D immersive interfaces that provide better semantic transparency to novices than experts.

4.3.8 Threats to Validity

The results of the online survey could have been more satisfactory. While analysing the results, we discovered more faulty study designs, primarily related to the uncontrolled variables.

- The criteria evaluation will be more accurate if we conduct it on every interface type with the same case study.
- The impact of expertise level on evaluation may have a different result if we have enough experts.
- The dual questions designed for complexity have a low Cronbach’s alpha value, which makes the results less reliable.
- The spatial relation result was concluded with only one pair of representations (two types of interfaces for one case study). The result will be more convincing if it is based on plural comparisons.

These new biases make the survey results less reliable but still provide valuable results.

4.4 Single-Person Controlled Experiment

4.4.1 Research Questions

The single-person controlled experiment aims to get a global vision of the 3D immersive MBSA interface performance. The participants will be university students. We intend to explore the following research questions: RQ1 (User Satisfaction), RQ2 (Visual Notation), RQ3(Presence), and RQ4 (Cognitive Load). RQ5 (Group Dynamics) is excluded in this experiment as it requires only one participant to complete the assigned tasks in the session.

Table 37 Validation methods and research questions (Expe_S)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

4.4.2 Single-Person Controlled Experiment Participants

The primary sources for the single-person controlled experiment participants will be the students from the University INP Grenoble. The experiment was conducted during their first-year practical course on systems engineering. Due to the time limits and the number of available devices, 72 students participated in the single-person controlled experiment. One of the main reasons we chose those students is that during their previous practical course that preceded the experiment, they had already been trained on the process's activities for designing engineered systems for 15 hours. A helicopter case study was used to introduce the activities. In contrast, an electric toothbrush served during four practical exercises (needs elicitation, system specification, system architecture, and a recursive process at the subsystems level). The modelling software Catia Magic with SysML as the modelling language was systematically used to encode data. Therefore, students are familiar with the architecture process and tools instead of being completely innocent in this domain. As a result, we saved time, as we did not need to invite the same students to achieve the experimental tasks using the conventional modelling software. We can directly experiment with the immersive 3D modelling interface and compare results with an evaluation of the 2D modelling interface achieved in the practical exercises of the course on Systems Engineering. 72 engineering students participated in the single-person controlled experiment.

4.4.3 Single-Person Controlled Experiment Design

As explained in section 4.4.2, the students will have already completed the exercise during the practical course. At the end of the last practical course, all students were asked to complete a questionnaire on user satisfaction, visual notation and cognitive load to evaluate the 2D diagrammatic MBSA interface. A second session gathered 72 students to assess the same validation criteria with the 3D immersive MBSA interface. Therefore, the independent variables are the modelling interface: 2D diagrammatic MBSA interface and 3D immersive MBSA interface. The dependent variables remain the validation criteria assessed through the questionnaire.

During the experiment, the participants had access to PowerPoint slides that illustrated the step-by-step process. The slides included the introduction of the system mission and the tasks to complete. Training using video tutorials applied to a telescope system enabled participants to learn how to use the 3D immersive MBSA interface. Finally, the slide deck included the exercise with the toothbrush system, followed by a questionnaire to fill out.

4.4.3.1 Setup

The room Mezzanine is reserved for the single-person controlled experiment. In the Mezzanine, there are four desktops with VR HMDs. Two other laptops are used at the same time. For each session of the single-person controlled experiment, we have six machines available for use. The participant can access PowerPoint slides on each machine describing the experiment process (slides in Appendix I). Students read the instructions at their own pace and proceed with the experiment. Camilo, one of the engineers in the laboratory, will be there to answer questions and play the role of the timekeeper. In front of each desk, there is a plastic-sealed reminder card. The card illustrates the buttons used in the 3D immersive MBSA interface. Two cards are made depending on the kind of VR HMD (HTC Vive Cosmos and Cosmos Elite).



Figure 127 Set up of a single-person controlled experiment

4.4.3.2 Case Study

The case study is a toothbrush case study. The electric toothbrush system has the mission of cleaning users' teeth. The electric toothbrush will also monitor brushing data to increase brushing efficiency by sending recommendations to the user through a mobile application. The electric toothbrush system will be connected to collect the following data: start time, end time, %

of time used for each mode, and time spent with a brushing pressure exceeding the threshold specified by healthcare recommendations.

The participant needs to develop the architecture of the Connected Electric Toothbrush System. The scope of the system analysis will be limited to the system level and pave the way to the definition of system requirements. During the experiment, they have four tasks to achieve:

- Define the life cycle of the Connected Electric Toothbrush System
- Define the external entities for a context of use (e.g., at home) of the Connected Electric Toothbrush System.
 - Stakeholders
 - External systems
- For a context of use (e.g., at home), define the external interfaces (ports and flows) between the Connected Electric Toothbrush System and the external entities.
- For a context of use (e.g., at home), define the Connected Electric Toothbrush System's functional scenario (functional chains).
 - Scenario: To brush teeth

4.4.3.3 Questionnaire Design

As we aim for research questions 1,2,3, and 4, the survey will include questions about user satisfaction, visual notation, presence, and cognitive load. A link to the online questionnaire is provided at the end of the slides provided to the participants.

4.4.3.4 Protocol (1h)

The experiment starts with filling out a consent form. Then, the participants read the introduction to the mission of an Electric Toothbrush System from the slides, followed by a list of system architecture tasks to be achieved. Next, they have a 15-minute training with the telescope system, after which they engage in a 25-minute exercise with the Electric Toothbrush System. Finally, participants complete a questionnaire that takes about 15 minutes to finish.

4.4.4 Single-Person Controlled Experiment Validation

A pilot study is conducted with the students before the large-scale experiment to identify design issues and evaluate the experiment's feasibility. Two researchers participated in this pilot study: one engineer and one professor from the laboratory. The case study was changed to a less complicated one to control the timing and became the toothbrush system. The videos were confirmed to explain the interaction with virtual environments clearly. The last task in the exercise is changed to an optional one in case the participant doesn't have enough time for the questionnaire. Some adjustments have been made to the controller button reminder cards for a more explicit demonstration.

4.4.5 Single-Person Controlled Experiment Analysis

The time gap between the students' practical course and the single-person controlled experiment is about a month. The students filled out one questionnaire online for the 2D diagrammatic MBSA interface right after their practical course. One month later, they participated in the single-person controlled experiment and completed another questionnaire for the 3D immersive MBSA interface. Considering the anonymous questionnaire's time gap, sample size, and challenging traceability, we used the independent sample t-test for the statistical analysis.

Research question 1 (i.e., User satisfaction) has several criteria. The SUS questionnaire assesses usability and an independent sample t-test will be conducted on the SUS score of 2D diagrammatic and 3D immersive MBSA interfaces. The rest of the criteria will be assessed separately in the same way. Moreover, a bar will illustrate the percentage of answers for the rest of the criteria.

For research question 2 (i.e., visual notation), the approach explained in 5.4 Online Survey will be applied here. Different aspects of the principles of PoN will be compared between the 2D diagrammatic MBSA interface and the 3D immersive MBSA interface with the independent sample t-test.

Research question 3 (i.e., presence) only concerns our case's 3D immersive MBSA interface. The score of different factors from IPQ will be calculated, and a diagram proposed by IPQ's official site showing the "presence profile" will be plotted. The diagram plans Spatial Presence (SP), Involvement (INV), and Experienced Realism (REAL) on three axes (range 0-7), and additionally, the General item (Presence) as a bow on the left.

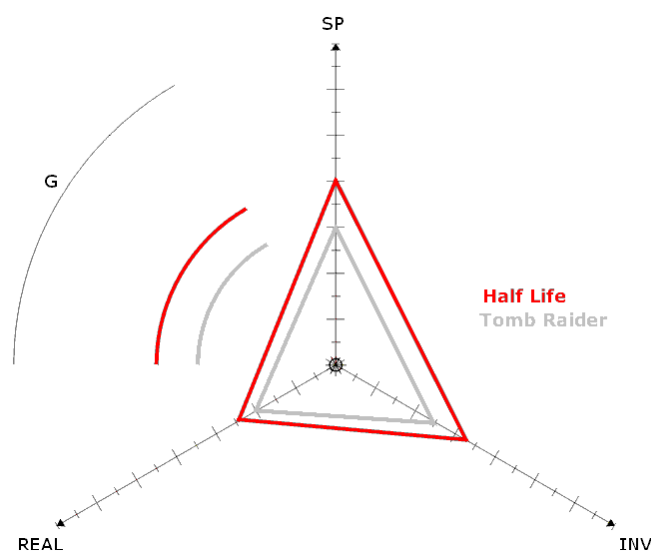


Figure 128 Example of "Presence profile" from study PQII

A scale proposed by (Melo et al. 2023) will be used to qualitatively interpret raw scores (G, SP, REAL, and INV). Qualitative grading encompasses the acceptability dimension and analogous academic grading scales ranging from A to F, as well as the adjective of such scores on a scale from Excellent to Unacceptable.

Meanwhile, on the official site of IPQ, researchers are allowed to upload their experiment data to their database, which enables other researchers to download these data from their database. Until the redaction of this manuscript, there are 542 cases in the database. The data is documented through several variables on researchers and studies. One of these is visual stimuli. The type 3D graphics (stereoscopic) via a head-mounted display (HMD) (coded as 6) is used to filter the data since it's the same type of visual stimuli with our 3D immersive MBSA interface. There are 37 studies belonging to this type. The means of IPQ raw scores from these 37 studies are calculated and compared with the 3D immersive MBSA interface.

For research, question 4 (i.e., cognitive load), an independent sample t-test (as the same reason for user satisfaction) will be conducted to compare the NASA-TLX raw score of the 2D diagrammatic and 3D immersive MBSA interfaces.

4.4.6 Single-Person Controlled Experiment Results

Independent sample t-test for research question 1:

RQ1: Does the 3D immersive MBSA interface increase **user satisfaction** for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface?

Table 38 Factors of independent sample t-test

Interface	Independent Variable
1	3D interface
0	2D interface

- **System Usability Scale (SUS)**

Table 39 Group statistics of independent sample t-test for SUS

	Interface	N	Mean	Std. Deviation	Std. Error Mean
SUS	1	55	55.6364	13.25688	1.78756
	0	75	41.4667	12.02540	1.38857

The 55 participants who had the exercise in the single-person controlled experiment with 3D immersive MBSA interface (M=55.64, SD=13.26) compared to the 75 participants who had the practical course with 2D diagrammatic MBSA interface (M=41.47, SD=12.03) demonstrated significantly better SUS score, $t(128)=6.260$, $p<.001$, $d=1.128$, representing a significant effect.

- **Perceived efficiency**

Table 40 Group statistics of independent sample t-test for perceived efficiency

	Interface	N	Mean	Std. Deviation	Std. Error Mean
PEfficiency	1	53	2.0755	1.01620	.13959
	0	76	2.3421	.91728	.10522

The 53 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=2.08, SD=1.02) compared to the 76 participants who had the practical course with 2D diagrammatic MBSA interface (M=2.34, SD=.92) demonstrated no significant difference of perceived efficiency, $t(127)=-1.554$, $p=.123$, $d=-.278$, representing a small effect.

- **Perceived effectiveness**

Table 41 Group statistics of independent sample t-test for perceived effectiveness

	Interface	N	Mean	Std. Deviation	Std. Error Mean
PEffectiveness	1	54	2.4259	.88172	.11999
	0	76	2.8553	.72487	.08315

The 54 participants who had the exercise in the single-person controlled experiment with 3D immersive MBSA interface (M=2.43, SD=.88) compared to the 76 participants who had the practical course with 2D diagrammatic MBSA interface (M=2.86, SD=.72) demonstrated significantly worse perceived effectiveness, $t(128)=-3.04$, $p=.003$, $d=-.541$, representing a medium effect.

- **Perceived retention rate**

Table 42 Group statistics of independent sample t-test for perceived retention rate

	Interface	N	Mean	Std. Deviation	Std. Error Mean
RR	1	54	2.6296	.91726	.12482
	0	75	2.4533	1.03053	.11899

The 54 participants who completed the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=2.63, SD=.92) compared to the 75 participants who completed the practical course with a 2D diagrammatic MBSA interface (M=2.45, SD=1.03) demonstrated no significant difference in perceived retention rate, $t(127)=1.003$, $p=.318$, $d=.179$, representing a small effect.

- **Intention to use in the future**

Table 43 Group statistics of independent sample t-test for intention to use in the future

	Interface	N	Mean	Std. Deviation	Std. Error Mean
ITU	1	54	2.1111	.96479	.13129
	0	73	2.5616	.89732	.10502

The 54 participants who had the exercise in the single-person controlled experiment with 3D immersive MBSA interface (M=2.11, SD=.96) compared to the 73 participants who had the

practical course with 2D diagrammatic MBSA interface (M=2.56, SD=.90) demonstrated significantly worse intention to use in the future, $t(125)=-2.709$, $p=.008$, $d=-.486$, representing a small to medium effect.

- **Usage fun**

Table 44 Group statistics of independent sample t-test for usage fun

	Interface	N	Mean	Std. Deviation	Std. Error Mean
UF	1	54	3.3148	.90750	.12349
	0	73	1.2192	1.03073	.12064

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.31, SD=.91) compared to the 73 participants who had the practical course with 2D diagrammatic MBSA interface (M=1.22, SD=1.03) demonstrated significantly better usage fun, $t(125)=11.909$, $p<.001$, $d=2.138$, representing a significant effect.

- **User motivation**

Table 45 Group statistics of independent sample t-test for user motivation

	Interface	N	Mean	Std. Deviation	Std. Error Mean
UM	1	54	2.7222	.95989	.13062
	0	72	1.7500	.96049	.11319

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=2.72, SD=.96) compared to the 72 participants who had the practical course with 2D diagrammatic MBSA interface (M=1.75, SD=.96) demonstrated significantly better user motivation, $t(124)=5.624$, $p<.001$, $d=1.012$, representing a large effect.

Independent sample t-test for research question 2:

RQ2: Does the 3D immersive MBSA interface provide better **visual notations** regarding the Physics of Notation (PoN) than the 2D diagrammatic MBSA interface?

- **Perceptual Discriminability**

Table 46 Group statistics of independent sample t-test for perceptual discriminability

	Interface	N	Mean	Std. Deviation	Std. Error Mean
PD	1	54	3.8704	.68807	.09363
	0	70	2.7786	.61170	.07311

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.87, SD=.69) compared to the 70 participants who had the practical course with 2D diagrammatic MBSA interface (M=2.78, SD=.61) demonstrated significantly better Perceptual Discriminability, $t(122)=9.332$, $p<.001$, $d=1.690$, representing a large effect.

- **Semantic Transparency**

Table 47 Group statistics of independent sample t-test for semantic transparency

	Interface	N	Mean	Std. Deviation	Std. Error Mean
ST	1	54	3.4074	.61486	.08367
	0	70	2.7357	.69010	.08248

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.41, SD=.61) compared to the 70 participants who had the practical course with 2D diagrammatic MBSA interface (M=2.74, SD=.69) demonstrated significantly better Semantic Transparency, $t(122)=9.332$, $p<.001$, $d=1.020$, representing a large effect.

- **Complexity Management**

Table 48 Group statistics of independent sample t-test for complexity management

	Interface	N	Mean	Std. Deviation	Std. Error Mean
CM	1	54	3.7037	.79217	.10780
	0	69	3.1449	.86220	.10380

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.70, SD=.79) compared to the 69 participants who had the practical course with 2D diagrammatic MBSA interface (M=3.14, SD=.86) demonstrated significantly better Complexity Management, $t(121)=3.695$, $p<.001$, $d=.671$, representing a medium effect.

- **Visual Expressiveness**

Table 49 Group statistics of independent sample t-test for visual expressiveness

	Interface	N	Mean	Std. Deviation	Std. Error Mean
VE	1	54	3.5741	.66167	.09004
	0	69	2.5072	.65019	.07827

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.57, SD=.66) compared to the 69 participants who had the practical course with 2D diagrammatic MBSA interface (M=2.51, SD=.65) demonstrated significantly better Visual Expressiveness, $t(121)=8.961$, $p<.001$, $d=1.628$, representing a large effect.

- **Graphic Economy**

Table 50 Group statistics of independent sample t-test for graphic economy

	Interface	N	Mean	Std. Deviation	Std. Error Mean
GEs	1	54	3.9259	.92862	.12637
	0	68	3.8235	.92947	.11271

The 54 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=3.93, SD=.93) compared to the 68 participants who had the practical course with a 2D diagrammatic MBSA interface (M=3.82, SD=.93) demonstrated no significant difference of Graphic Economy, $t(120)=.605$, $p=.547$, $d=.110$, representing a small effect.

IPQ Raw Score for Research Question 3:

RQ3: Does a 3D immersive MBSA interface possess a sufficient level of **presence** so that the users can stay focused on their tasks while in the virtual world?

Table 51 Descriptive statistics for four factors of IPQ

	N	Minimum	Maximum	Mean	Std. Deviation
Presence	52	.00	6.00	3.1731	1.55578
Spatial Presence	38	1.60	5.00	3.5053	.83664
Involvement	41	1.00	5.00	2.9878	1.10107
Experienced Realism	36	.25	4.25	2.2083	.94962

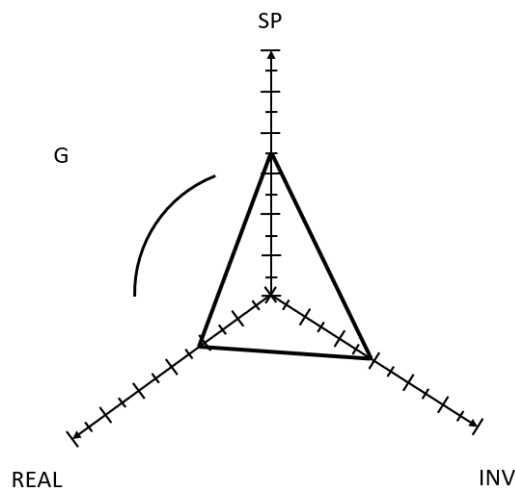


Figure 129 IPQ result

The following figure illustrates the IPQ scores of the 3D immersive MBSA interface and the means of IPQ scores from 37 studies with 3D graphics (stereoscopic) visa HMD.

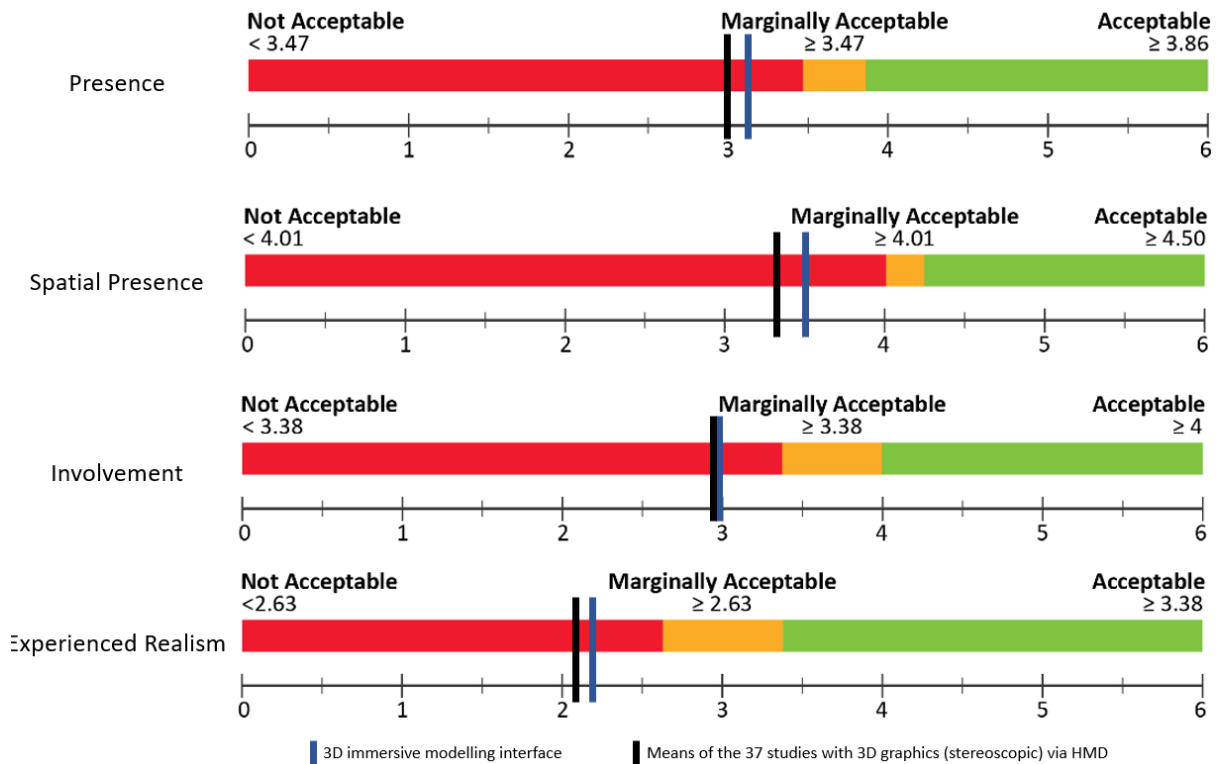


Figure 130 IPQ scores (3D immersive MBSA interface scores and mean scores)

Independent sample t-test for research question 4:

RQ4: Does a 3D immersive MBSA interface reduce the cognitive load for co-designing a system architecture compared with a 2D diagrammatic MBSA interface?

NASA-TLX raw score is calculated and compared between the 2D diagrammatic MBSA interface and the 3D immersive MBSA interface.

Table 52 Group statistics of independent sample t-test for NASA-TLX raw score

	Interface	N	Mean	Std. Deviation	Std. Error Mean
TLXRawScore	1	40	39.2083	14.34336	2.26788
	0	47	53.5638	16.11281	2.35029

The 40 participants who had the exercise in the single-person controlled experiment with a 3D immersive MBSA interface (M=39.20, SD=14.34) compared to the 47 participants who had the practical course with 2D diagrammatic MBSA interface (M=53.56, SD=16.11) demonstrated significantly lower task load, $t(85)=-4.354$, $p<.001$, $d=-.937$, representing a large effect.

4.4.7 Single-Person Controlled Experiment Discussion

The 3D immersive MBSA interface returns a significantly better SUS score than the 2D diagrammatic MBSA interface from the independent sample t-test. However, regarding User Satisfaction, several criteria for the 3D immersive MBSA interface demonstrate a lower

performance than the 2D immersive MBSA interface. The perceived efficiency of a 3D immersive MBSA interface is slightly worse than that of a 2D diagrammatic MBSA interface. The perceived effectiveness and intention to use 3D immersive MBSA interfaces in the future are significantly worse than 2D diagrammatic MBSA interfaces. This relates to the flawed design (on interaction mostly) and the hardware's limits. The interaction with the environment can sometimes be confusing, especially for the VR novices. It becomes even worse when the exercise is interrupted by some connection problem. The user needs to wait to re-establish the connection or restart the project. It could lead to a low perceived efficiency and effectiveness. At the end of the questionnaire, the participants can leave general feedback or some remarks they noticed during the experiment. Since it was optional, there were only a few, but some mentioned VR's illness and the helmet's discomfort. Those factors prevent the users from having a decent VR experience. However, from the general feedback on user satisfaction, the 3D immersive MBSA interface demonstrates some potential values of research in this direction.

After the exercises in both modelling interfaces, the participants felt that the 3D immersive MBSA interface had better visual notation than the 2D diagrammatic MBSA interface, significantly different on four criteria. This confirmed our alternative hypothesis of research question 2. As for the cognitive load, the 3D immersive MBSA interface reduces the cognitive load more than the 2D diagrammatic MBSA interface. This ensures the interest in research with the 3D immersive MBSA interface.

The result of the Presence evaluation could be more satisfactory based on the scale proposed by Melo. According to the dataset on the IPQ website, the average presence score is slightly lower than the 3D immersive MBSA interface. It reflects the limitations of current technology, both at the software and hardware level. Furthermore, importing incredibly realistic models into the environment will surpass the software's capacity and overload the machine. The tradeoff between the actual level of models and the overall performance of the 3D immersive MBSA interface is another point that should be examined. In any case, future work with more intuitive interaction design and better 3D models in VR environments can improve feelings of presence, which can be booted exponentially by the development of VR technology.

4.5 Multi-Person Controlled Experiment

4.5.1 Research Questions

In this section, the multi-person controlled experiment is presented. The research questions focused on multi-person controlled experiments will be research questions 1,2,4,5. For research questions 1,2,4, the measurements will be similar to the single-person controlled experiment with a difference in working context. Instead of individual exercises, the focus here is on group exercises.

Table 53 Validation methods and research questions (Expe_M)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

4.5.2 Multi-Person Controlled Experiment Participants

The participants are always the students from the University INP Grenoble. As the multi-person controlled experiments are anticipated to take slightly longer than single-person controlled experiments and require multiple students to be available simultaneously, 48 students were eventually able to participate.

4.5.3 Multi-Person Controlled Experiment Design

The experiment will take about an hour and a half. The students will have two group work sessions, one (25') for a 2D diagrammatic MBSA interface and another for a 3D immersive MBSA interface. Only one case study was used in the multi-person controlled experiment. The order of the two modelling interfaces is counterbalanced to minimise the potential learning effect as much as possible. The students will partially co-develop the architecture of a system. They will be sitting before the screen wall that displays the modelling interface. One animateur on the corner will interact with the modelling interface.

Like the single-person controlled experiment, the modelling interface—2D diagrammatic and 3D immersive MBSA interfaces—will be the independent variable. The criteria assessed through the questionnaire will be the dependent variable.

4.5.3.1 Setup

The room Vision-R is reserved for the multi-person controlled experiment. One desktop in Vision R with VR HMD is used. For each session of the multi-person controlled experiment, 4-6 participants are sitting in front of the Powerwall. Participants can access PowerPoint slides on the Powerwall screen (slides in Appendix J). The slides will help represent the experiment's protocol, the system's development mission, and the tasks to complete. The slides also contain two short introductory videos about the system to create. I will operate the desktop and communicate with other participants to co-develop the system.

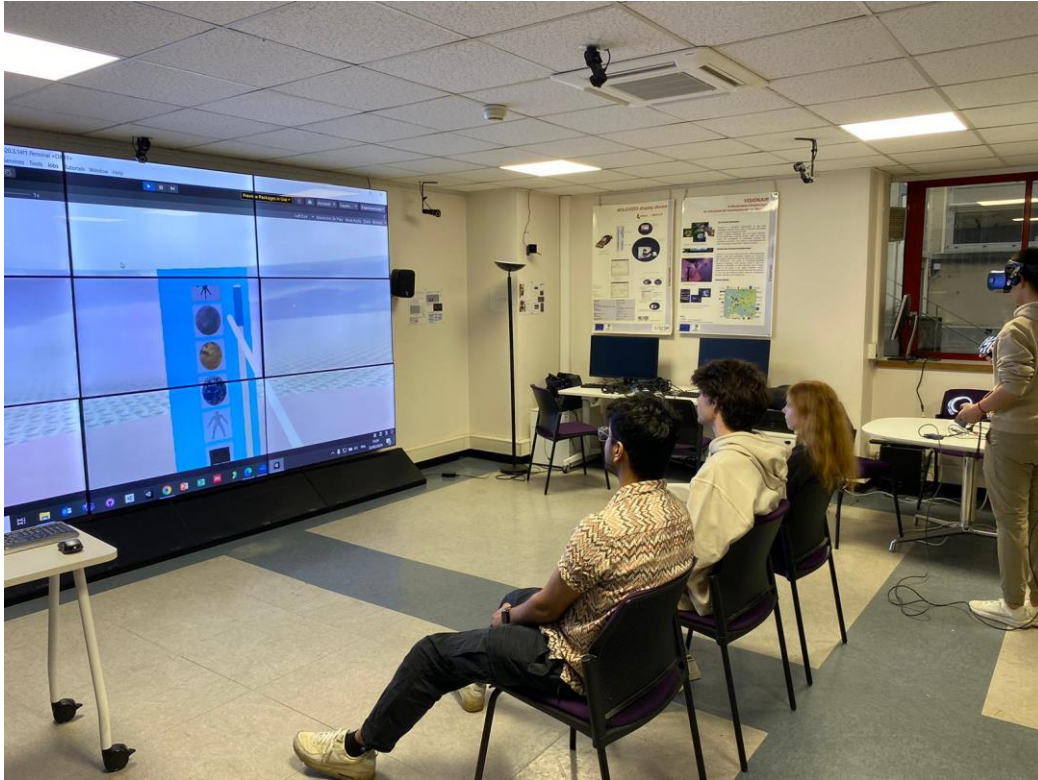


Figure 131 Set up of a multi-person controlled experiment

4.5.3.2 Case Study

The case study in a multi-person controlled experiment is the wearable on-body drug delivery system from Becton Dickinson, a leading company developing medical products. The wearable on-body connected drug delivery system is an automated single-use injection device, filled at the hospital, to automatically administer 2-10 mL of a drug to cure chronic diseases and with frequent or higher autoimmune diseases and immuno-oncology drug viscosity within 2-10 minutes. It must also record injection data (device ID, injection start time, and injection end time) to document the correct completion of the drug delivery process, that is, detect deviation from the prescribed injection protocol without modifying the user injection experience.

The participant needs to develop the architecture of the wearable on-body drug delivery system. During the collaborative session, they will achieve the following four tasks in both modelling interfaces:

- Define the life cycle of the wearable on-body drug delivery system
- Define the external entities for a context of use (e.g., at home) of the wearable on-body drug delivery system.
 - Stakeholders
 - External systems
- For a specific context (e.g., use at home), define the external interfaces (ports and flows) between the wearable on-body drug delivery system and the external entities.

- For a context of use (e.g., use at home), define the wearable on-body drug delivery system's functional scenario (functional chains).
 - Scenario: To inject a drug dose

4.5.3.3 Survey Design

The survey will have questions about user satisfaction, visual notation, cognitive load, and group dynamics for research questions 1, 2, 4, and 5. The questions about research questions 1,2 and 4 are the same as in the single-person controlled experiment, with the difference in the collaborative work context. Only one question is involved for research question 5 since the rest of the criteria are evaluated with the facts gathered from the experiment data.

4.5.3.4 Protocol (1.5h)

The experiment starts with filling out a consent form. Then, an overview of the wearable on-body drug delivery system with two videos will be presented to participants, followed by an introduction to the mission of a new wearable on-body drug delivery system and a list of system architecture tasks to complete. Next, they have a 25-minute exercise with a wearable on-body drug delivery system in a 2D diagrammatic MBSA interface and another 25-minute exercise in a 3D immersive MBSA interface with a counterbalanced order). Finally, participants complete a questionnaire that takes about 20 minutes to finish.

4.5.4 Multi-Person Controlled Experiment Validation

Before the period of this multi-person controlled experiment, we had already begun another type of experiment: formative usability tests with MBSA experts. Due to the similarity of the multi-person controlled experiment with the single-person controlled experiment and the formative usability tests, we did not conduct a pilot study for this experiment.

Unfortunately, during the first day of the multi-person controlled experiment, we encountered some issues mainly related to research question 5:

- Throughout the experiment, despite our encouragement for participants to engage in think-aloud activities or communicate with other group members, there was minimal verbal interaction during the entire session. Most groups exhibited a passive communication performance, a factor we did not anticipate. This unexpected phenomenon rendered the collection of factors related to research question 5 unfeasible. Consequently, we decided to introduce additional questions in the survey to substitute for the previously intended factors. As a result, the final evaluation criteria from the survey encompass perceived efficiency of communication, perceived effectiveness of communication, and member activeness.

- We observed that participants, once they completed tasks in one interface, were unwilling to repeat the same tasks in the other interface, regardless of which modelling interface they started with. This reluctance further exacerbated the participants' low motivation to engage in discussion. Therefore, rather than requiring them to complete all tasks in both modelling interfaces, they will now perform half of the functions in one of the two interfaces and complete the remaining tasks with the other.

4.5.5 Multi-Person Controlled Experiment Analysis

The analysis method will be similar to the technique in a single-person controlled experiment for questions 1,2 and 4. The process is the same for question 5 after the adjustment since it is based on the 5-point Likert scale questions like the other group of criteria. Since we collected the evaluation for the 2D and 3D simultaneously, with limited participants, we decided to use a paired sample t-test for the multi-person controlled experiment.

4.5.6 Multi-Person Controlled Experiment Results

Paired sample t-test for research question 1:

RQ1: Does the 3D immersive MBSA interface increase **user satisfaction** for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface?

- **System Usability Scale (SUS)**

Table 54 Paired samples statistics of paired sample t-test for SUS

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SUS2D	50.4545	44	14.34898	2.16319
	SUS3D	56.6477	44	14.19285	2.13965

The results indicated that there was no significant difference between the SUS score of the 2D diagrammatic MBSA interface (M=50.45, SD=14.35) and the SUS score of the 3D immersive MBSA interface (M=56.65, SD=14.19), $t(43)=-1.893$, $p=.065$, $d=-.285$, representing a small effect.

- **Perceived efficiency**

Table 55 Paired samples statistics of paired sample t-test for perceived efficiency

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Efficient2D	2.2391	46	.89901	.13255
	Efficient3D	2.0000	46	1.15470	.17025

The results indicated that there was no significant difference between the perceived efficiency of the 2D diagrammatic MBSA interface (M=2.24, SD=.90) and the perceived efficiency of the 3D immersive MBSA interface (M=2.00, SD=1.15), $t(45)=.968$, $p=.338$, $d=.149$, representing a small effect.

- **Perceived effectiveness**

Table 56 Paired samples statistics of paired sample t-test for perceived effectiveness

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Effectiveness2D	2.8913	46	.84927	.12522
	Effectiveness3D	2.6087	46	.99952	.14737

The results indicated that there was a significant difference between the perceived effectiveness of the 2D diagrammatic MBSA interface (M=2.89, SD=.85) and the perceived effectiveness of the 3D immersive MBSA interface (M=2.61, SD=1.00), $t(45)=1.642$, $p=.108$, $d=.242$, representing a small effect.

- **Perceived retention rate**

Table 57 Paired samples statistics of paired sample t-test for perceived retention rate

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	RR2D	2.3913	46	.95402	.14066
	RR3D	2.9130	46	.96208	.14185

The results indicated that there was a significant difference between the perceived retention rate of the 2D diagrammatic MBSA interface (M=2.39, SD=.95) and the perceived retention rate of the 3D immersive MBSA interface (M=2.91, SD=.96), $t(45)=-2.482$, $p=.017$, $d=-.366$, representing a small effect.

- **Intention to use in the future**

Table 58 Paired samples statistics of paired sample t-test for intention to use in the future

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	ITU2D	2.6304	46	.85267	.12572
	ITU3D	2.1087	46	1.12008	.16515

The results indicated that there was a significant difference between the intention to use in the future of 2D diagrammatic MBSA interface (M=2.63, SD=.85) and the intention to use in the future of 3D immersive MBSA interface (M=2.11, SD=1.12), $t(45)=2.228$, $p=.031$, $d=.329$, representing a small effect.

- **Usage fun**

Table 59 Paired samples statistics of paired sample t-test for usage fun

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	UF2D	1.4348	46	1.10860	.16345
	UF3D	2.8913	46	.84927	.12522

The results indicated that there was a significant difference between the intention to use in the future of 2D diagrammatic MBSA interface (M=1.43, SD=1.11) and the intention to use in

the future of 3D immersive MBSA interface (M=2.89, SD=.85), $t(45)=-7.256$, $p<.001$, $d=-1.07$, representing a significant effect.

- **User motivation**

Table 60 Paired samples statistics of paired sample t-test for user motivation

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	UM2D	1.7111	45	1.03621	.15447
	UM3D	2.4667	45	1.01354	.15109

The results indicated that there was a significant difference between the intention to use a 2D diagrammatic MBSA interface (M=1.71, SD=1.04) and the intention to use a 3D immersive MBSA interface (M=2.47, SD=1.01) in the future, $t(44)=-3.04$, $p=.004$, $d=-.453$, representing a small effect.

Paired sample t-test for research question 2:

RQ2: Does the 3D immersive MBSA interface provide better **visual notations** regarding the Physics of Notation (PoN) than the 2D diagrammatic MBSA interface?

- **Perceptual Discriminability**

Table 61 Paired samples statistics of paired sample t-test for perceptual discriminability

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PD2D	2.7889	45	.94441	.14079
	PD3D	3.8778	45	.79169	.11802

The results indicated that there was a significant difference between the Perceptual Discriminability of the 2D diagrammatic MBSA interface (M=2.79, SD=.94) and the Perceptual Discriminability of the 3D immersive MBSA interface (M=3.88, SD=.79), $t(44)=-5.744$, $p<.001$, $d=-.856$, representing a significant effect.

- **Semantic Transparency**

Table 62 Paired samples statistics of paired sample t-test for semantic transparency

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	ST2D	2.7065	46	.74964	.11053
	ST3D	3.9674	46	.66167	.09756

The results indicated that there was a significant difference between the Semantic Transparency of the 2D diagrammatic MBSA interface (M=2.71, SD=.75) and the Semantic Transparency of the 3D immersive MBSA interface (M=3.97, SD=.66), $t(45)=-8.379$, $p<.001$, $d=-1.235$, representing a significant effect.

- **Complexity Management**

Table 63 Paired samples statistics of paired sample t-test for complexity management

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	CM2D	3.3913	46	.80217	.11827
	CM3D	3.9348	46	.90436	.13334

The results indicated that there was a significant difference between the Complexity Management of the 2D diagrammatic MBSA interface (M=3.39, SD=.80) and the Complexity Management of the 3D immersive MBSA interface (M=3.93, SD=.90), $t(45)=-3.449$, $p<.001$, $d=-.508$, representing a medium effect.

- **Visual Expressiveness**

Table 64 Paired samples statistics of paired sample t-test for visual expressiveness

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VE2D	2.8043	46	.83319	.12285
	VE3D	4.0000	46	.64979	.09581

The results indicated that there was a significant difference between the Visual Expressiveness of the 2D diagrammatic MBSA interface (M=2.80, SD=.83) and the Visual Expressiveness of the 3D immersive MBSA interface (M=4.00, SD=.65), $t(45)=-7.675$, $p<.001$, $d=-1.132$, representing a significant effect.

- **Graphic Economy**

Table 65 Paired samples statistics of paired sample t-test for graphic economy

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	GE2D	3.3778	45	1.09314	.16296
	GE3D	3.8667	45	.75679	.11282

The results indicated that there was a significant difference between the Graphic Economy of the 2D diagrammatic MBSA interface (M=3.38, SD=1.09) and the Graphic Economy of the 3D immersive MBSA interface (M=3.87, SD=.76), $t(44)=-2.875$, $p=.006$, $d=-.429$, representing a small effect.

Paired sample t-test for research question 4:

RQ4: Does a 3D immersive MBSA interface reduce the cognitive load for co-designing a system architecture compared with a 2D diagrammatic MBSA interface?

Table 66 Paired samples statistics of paired sample t-test for NASA-TLX raw score

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TLX2D	47.9012	27	14.91316	2.87004
	TLX3D	47.0370	27	14.06702	2.70720

The results indicated that there was no significant difference between the task load of the 2D diagrammatic MBSA interface (M=47.90, SD=14.91) and the task load of the 3D immersive MBSA interface (M=47.03, SD=14.07), $t(26)=-.296$, $p=.769$, $d=.057$, representing a small effect.

Paired sample t-test for research question 5:

RQ5: Does a 3D immersive MBSA interface increase the group dynamics for co-designing system architecture activities compared with a 2D diagrammatic MBSA interface?

- **Perceived efficiency of communication**

Table 67 Paired samples statistics of paired sample t-test for perceived efficiency of communication

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PEfficiency2D	2.8444	45	.82450	.12291
	PEfficiency3D	2.8889	45	.85870	.12801

The results indicated that there was no significant difference between the perceived efficiency of communication of the 2D diagrammatic MBSA interface (M=2.84, SD=.82) and the perceived efficiency of communication of the 3D immersive MBSA interface (M=2.89, SD=.86), $t(44)=-.467$, $p=.643$, $d=-.070$, representing a small effect.

- **Perceived effectiveness of communication**

Table 68 Paired samples statistics of paired sample t-test for perceived effectiveness of communication

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PEffectiveness2D	2.8222	45	.74739	.11141
	PEffectiveness3D	2.9556	45	.73718	.10989

The results indicated that there was no significant difference between the perceived effectiveness of communication of the 2D diagrammatic MBSA interface (M=2.82, SD=.75) and the perceived effectiveness of communication of the 3D immersive MBSA interface (M=2.96, SD=.74), $t(44)=-1.182$, $p=.244$, $d=-.176$, representing a small effect.

- **Member activeness**

Table 69 Paired samples statistics of paired sample t-test for member activeness

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	MA2D	2.9565	46	.59466	.08768
	MA3D	2.8696	46	.85916	.12668

The results indicated that there was no significant difference between the member activeness of the 2D diagrammatic MBSA interface (M=2.96, SD=.59) and the member activeness of the 3D immersive MBSA interface (M=2.87, SD=.86), $t(45)=.813$, $p=.420$, $d=.120$, representing a small effect.

4.5.7 Multi-Person Controlled Experiment Discussion

In the collaborative work context, the 3D immersive MBSA interface has a significantly better SUS score, perceived retention rate, usage fun, user motivation, and visual notations than the 2D diagrammatic MBSA interface. This indicates that the 3D immersive MBSA interface can be a solution to increase collaborative work performance.

Usually, better fun usage and user motivation will enhance the group dynamics. However, the three criteria of group dynamics are similar. An insufficient evaluation approach may cause this. Group dynamics are strongly related to the personalities of the participants. Because most of the group communicated barely, we could not get decent feedback even if we changed the approach to some questions in the questionnaire.

We failed to reduce the task load with the 3D immersive MBSA interface, such as the group dynamics. This may be due to a similar reason as the failure with group dynamics. In a situation where participants are reluctant to engage in discussions, the task load is likely to be significantly poor, regardless of the modelling interface. Therefore, a change in the modelling interface may not substantially impact the task load.

4.6 Formative Usability Tests with MBSA Experts

From the creator's perspective, the modelling interface only works if it is applied in a practical situation. The results obtained from the controlled experiments are particularly interesting for getting sound empirical evidence regarding the potential of using VR technologies for MBSE. Nevertheless, new contributions (theory, method, software...) in engineering design are, in practice, evaluated by expert approval; that is, experts use the results of the research – here is a new modelling interface for MBSA – and are invited to comment. This section details how this exploratory research used formative usability testing with our fully functional virtual prototype. These formative usability tests are more informal, with observations of MBSA experts and more communication between test moderators and MBSA experts, with the primary goal of evaluating how the MBSA experts perceive the HCI rather than how well they complete their tasks. At the end of this testing phase, we expect to clarify the requirements and design of the virtual environment to better satisfy MBSA experts' needs.

4.6.1 Research Questions

The formative usability tests will address research questions 1, 2, 3, and 4.

Table 70 Validation methods and research questions (Formative)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

4.6.2 Formative Usability Testing Participants

For the expert part, we were able to have some practitioners of system architecture. They are from industries that involve system engineering activities. For instance, we have some employees from Schneider Electric (SE). SE is a French multinational company that operates in the energy management and automation industry, specialising in products and solutions that enable energy efficiency and sustainability. They are familiar with system engineering activities due to their routine work, which includes design, integration, and system management. With experience using our 3D immersive MBSA interface, they can give us practical advice combined with their work experience. There will be three groups of two practitioners from 3 industrial companies.

- **Schneider Electric**

The first couple of participants are engineers developing new electric power systems at Schneider Electric. Both are MBSA methods & tools experts. Participant A is an MBSA expert. Participant B is an MBSA practitioner. Both are novices in virtual reality. The 2D modelling interface they are evaluating is the custom SysML-based framework they use to architect digital automation and energy management systems with Catia Magic as modelling software.

- **Becton Dickinson**

The second couple of participants are engineers developing advanced drug delivery systems at Becton Dickinson. Participant A is an expert on system architecture but not on MBSA. Participant B is an MBSA expert with less domain knowledge of drug delivery systems. Both are novices in virtual reality. Participant B's 2D diagrammatic MBSA interface is Catia Magic, and Participant A uses Word, Excel, and PowerPoint.

- **Axone**

The third couple of participants are engineers from Axone, a major actor in systems engineering in the nuclear field, specialising in the architectural approach to complex systems. Participant A and Participant B are both experts on MBSA and system architecture. The 2D

diagrammatic MBSA interface for Participant A and Participant B is the conventional SysML-based and Capella software for MBSA.

Table 71 Group with expertise level

		MBSA	System Architecture	Virtual Reality
Test 1 (SE)	Participant A	Expert	Expert	Novice
	Participant B	Expert	Practitioner	Novice
Test 2 (BD)	Participant A	Medium	Expert	Novice
	Participant B	Expert	Expert	Novice
Test 3 (Axone)	Participant A	Expert	Expert	Novice
	Participant B	Expert	Expert	Novice

4.6.3 Formative Usability Testing Design

We conducted three formative testing sessions with three groups of participants from three different industrial disciplines. Each session was scheduled for three hours, with an additional 20-minute break. Within each session, two moderators were present: moderator A was responsible for co-developing the system with the participants, and moderator B was responsible for interacting with the interface. The testing setup involved moderator A and two participants seated at the screen wall. Moderator B was positioned in the corner with a laptop and VR gear.

This validation method excludes the 2D diagrammatic MBSA interface, as participants are MBSA practitioners from the industrial sector who already possess experience with the 2D diagrammatic MBSA interface in their daily work. Only the 3D immersive MBSA interface is used for multi-person and single-person formative testing. The formative testing begins with multi-person formative testing, allowing participants to gain an initial understanding of the 3D immersive MBSA interface. They can discover the method and environment (visual metaphors, interaction metaphor, etc.) as a contributor passively involved. It will help them with the later single-person formative testing, where they actively engage with the VR environment themselves. Once they finish the multi-person formative testing, participants will proceed to the single-person formative testing, followed by a questionnaire and an interview.

4.6.3.1 Setup

The room Vision-R is reserved for the multi-person controlled experiment. One desktop in Vision R and one laptop with VR HMDs are used. For each session of the formative usability test, there is a multi-person formative user testing session (involving one desktop, similar to a multi-person controlled experiment) and a single-person formative user testing session (involving one desktop and one lap, as there are two experts in every group, similar to a single-person controlled experiment). A powerwall is used in the multi-person formative user testing session to share the view of the 3D immersive MBSA interface. At the same time, moderator

A will co-develop the system with experts, and moderator B will operate the desktop. Experts can access PowerPoint slides on the screen of Powerwall (in a multi-person formative user testing session) or the screen of a desktop/laptop (in a single-person formative user testing session) (slides in Appendix K).



Figure 132 Photo of formative usability testing

4.6.3.2 Case Studies

The two case studies used here for the multi-person and single-person formative studies are the same as the single-person and multi-person controlled experiments. The wearable on-body drug delivery system will be used for the multi-person formative study, and the electric toothbrush system will be used for the single-person formative study.

4.6.3.3 Post-formative Usability Testing Questionnaire

As we aim for research questions 1,2,3, and 4, the questionnaire will include questions about user satisfaction, visual notation, presence, and cognitive load. At the beginning of the questionnaire, demographic questions will be used in the online survey. The questions (except those for research question 3, i.e., the questions of IPQ) will be asked for both 2D diagrammatic and 3D immersive MBSA interfaces. Depending on every participant, the 2D diagrammatic MBSA they use may be conventional MBSE software like CATIA Magic or Word/Excel. The questionnaire is printed on paper since only 6 participants are involved.

4.6.3.4 Post-formative Usability Testing One-to-one Interview

We use interviews to obtain more detailed and thorough information on our 3D immersive MBSA interface that might be gleaned from the usability testing questionnaire. The interview

includes questions about the current 2D diagrammatic MBSA interface they use in the work, the 3D immersive MBSA interface they tested during the experiment, and their perspectives on the future of the 3D immersive MBSA interface development.

4.6.3.5 Protocol (3h)

The experiment starts with an introduction to the research motivation and objectives, a formative usability testing protocol, and a consent form (15 minutes). Then, participants engage in the multi-person formative user testing session for about 45 minutes. In this session, they will have an overview of a Wearable On-body Drug Delivery System (5 minutes), an introduction to the mission of a new Wearable On-body Drug Delivery System (5 minutes), a list of system architecture tasks to achieve (5 minutes), an exercise (25 minutes), and a quick overview of future modelling capabilities (new context, requirements, subsystems) (5 minutes). They have a 10-minute break after this session. Next, participants engage in the single-person formative user testing session for about 60 minutes. In this session, they will have an introduction to the mission of a new connected Electric Toothbrush System (3 minutes), a list of system architecture tasks to achieve (2 minutes), training (20 minutes), and an exercise (35 minutes). Then, they have another 10-minute break before they fill out a questionnaire, which is about 15 minutes, and have an interview for about 45 minutes.

4.6.4 Formative Usability Testing Validation

A pilot study is conducted before the formative study to identify design issues and evaluate the experiment's feasibility. Two researchers participated in this pilot study: one engineer and one professor from the laboratory.

4.6.5 Formative Usability Testing Analysis

The results of the questionnaire and interview will be analysed for each participant. Since it's a formative study with no statistical analysis, we will analyse the results based on each participant's experience. Finally, we will summarise all the feedback (questionnaire and interview) we collected.

4.6.6 Formative Usability Testing Results

Results for research question 1:

RQ1: Does the 3D immersive MBSA interface increase user satisfaction for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface?

The following figures represent the results of the evaluation for research question 1.

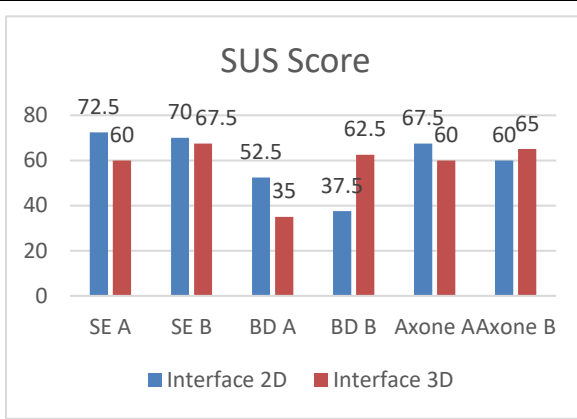


Figure 133 SUS scores

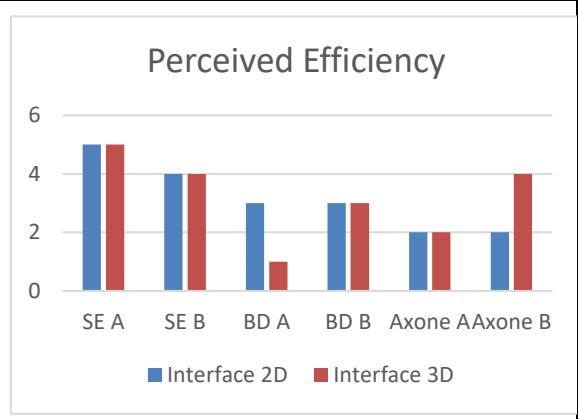


Figure 134 Perceived efficiency

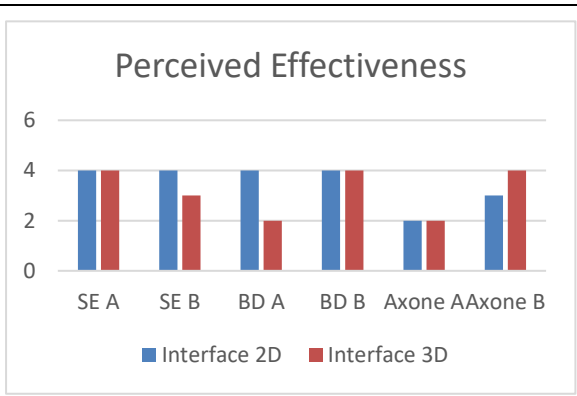


Figure 135 Perceived effectiveness

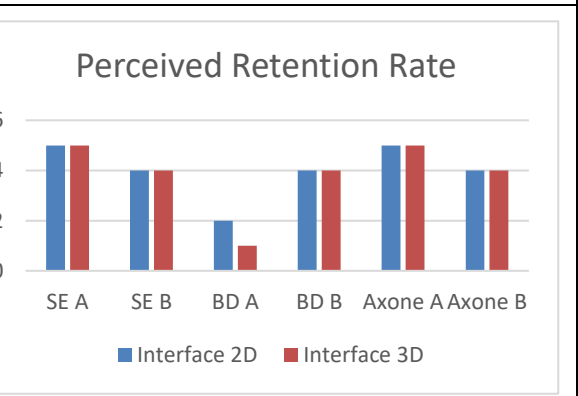


Figure 136 Perceived retention rate

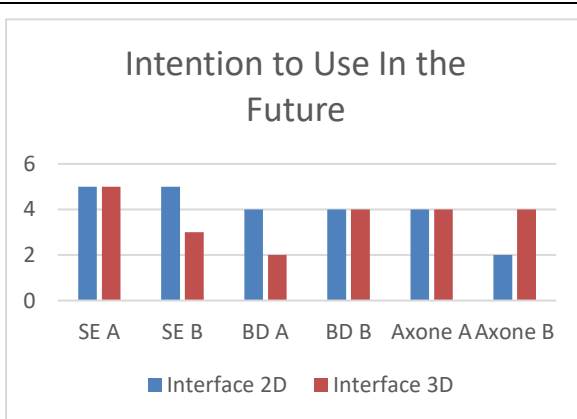


Figure 137 Intention to use in the future

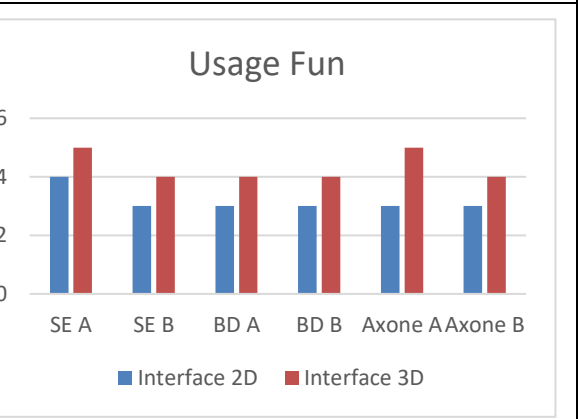


Figure 138 Usage fun

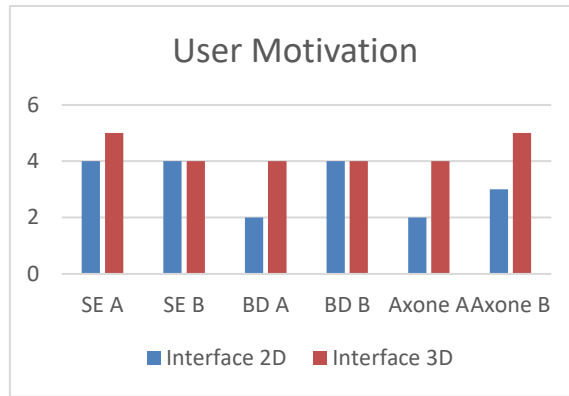


Figure 139 User motivation

Results for research question 2:

RQ2: Does the 3D immersive MBSA interface provide better visual notations regarding the Physics of Notation (PoN) than the 2D diagrammatic MBSA interface?

The following figures represent the results of the evaluation for research question 2.

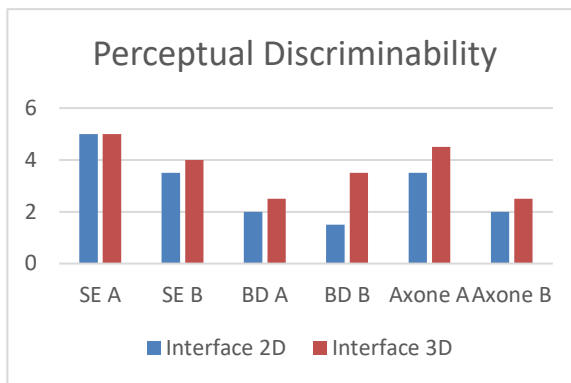


Figure 140 Perceptual discriminability

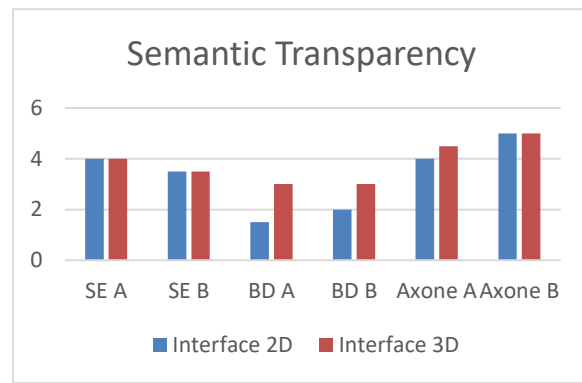


Figure 141 Semantic transparency

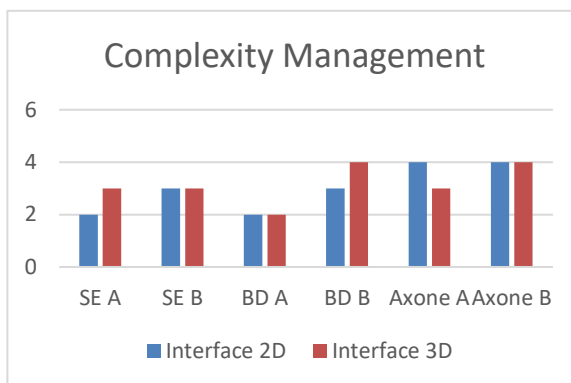


Figure 142 Complexity management

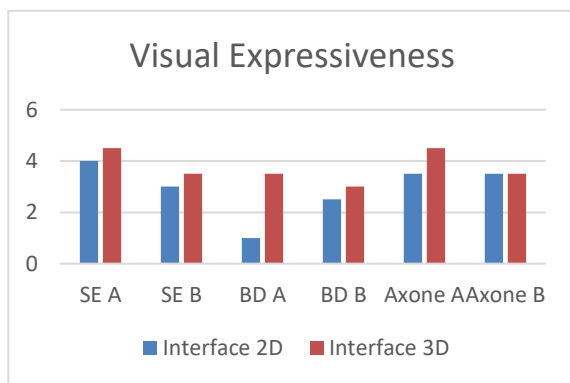


Figure 143 Visual expressiveness

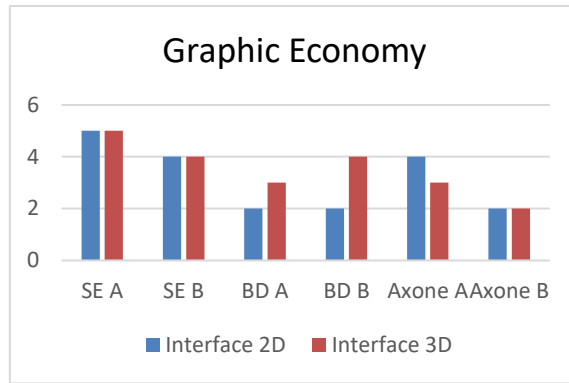


Figure 144 Graphic economy

Results for research question 3:

RQ3: Does a 3D immersive MBSA interface possess a sufficient level of **presence** so that the users can stay focused on their tasks while in the virtual world?

The following figures represent the results of the evaluation with IPQ for research question 3, with the standardised scale from Melo (Melo et al., 2023).

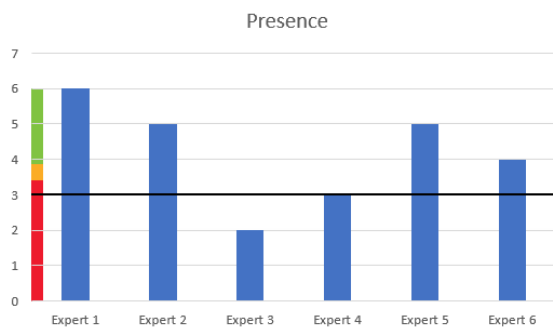


Figure 145 Presence

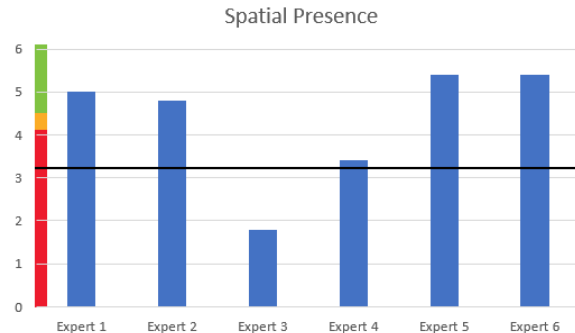


Figure 146 Spatial presence

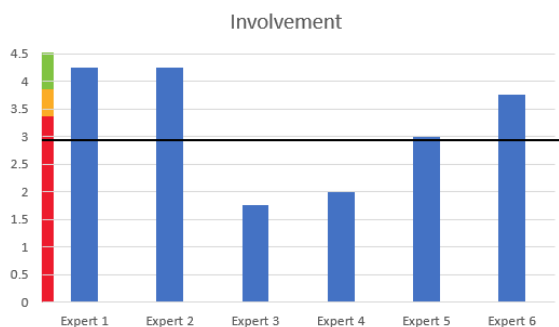


Figure 147 Involvement



Figure 148 Experienced realism

Results for research question 4:

RQ4: Does a 3D immersive MBSA interface reduce the cognitive load for co-designing a system architecture compared with a 2D diagrammatic MBSA interface?

The following figures represent the evaluation results with the NASA-TLX questionnaire for research question 4.

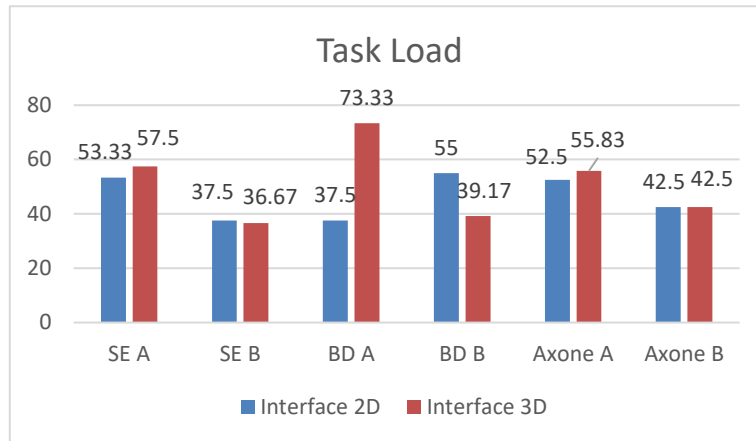


Figure 149 NASA-TLX raw score

Results of the interview:

- **Schneider Electric**

The overall evaluation of the user experience with the 3D immersive MBSA interface was very positive. They think the interface is straightforward to learn and use. They do not need training to use the interface, and there are very few buttons they need to remember to be able to use it. Participant A is very interested in integrating the 3D immersive MBSA interface with their framework to see how to create a model for their project. Furthermore, Participant A reported that the 3D immersive MBSA interface does not explore and use the full advantages and potentials of 3D space.

Based on their feedback, one aspect that we could have achieved with our interface, as we intended to achieve, was the attempt to present more intuitive visual notations with 3D models than 2D diagrammatic models. Moreover, even though direct representation, like images, can facilitate communication, it concerns them that the representation's level of abstraction may sometimes interfere with the final decision of the product.

Despite the brief case study and tasks, they can still sense the potential facility brought by the 3D immersive environment. As to the two interactive devices they used (HMD and power wall), participant B prefers the HMD, and Participant A thinks each has its advantages depending on the situation. For example, Powerwall is more suitable for cooperation. They both expect to integrate similar features into their work, but the real problem is how to achieve this integration properly. More importantly, what they hope is an incremental evolution instead of an instant and total revolution.

- **Becton Dickinson**

Participant B had to leave early and was absent for the interview section; thus, this part was conducted only with Participant A. Participant A still uses classic working software such as Word, PowerPoint, or Excel in their work routine. During the interview, the word “fun” was frequently mentioned by participant A. He thinks that the most distinctive aspect of the 3D immersive MBSA interface, compared to the conventional 2D diagrammatic MBSA interface, is the “fun” part. Using the 3D immersive MBSA interface occasionally lets them take a break from routine work. However, he has no intention of adopting a 3D immersive MBSA interface as their primary approach, given their familiarity with traditional 2D diagrammatic MBSA interface, and he expressed interest in combining our 3D immersive MBSA interface with current applied software.

According to Participant A, teamwork with a 3D immersive MBSA interface with the Powerwall can enhance communication. The 3D models provide more context than the 2D diagrammatic MBSA interface. Furthermore, the dependencies of models with PowerPoint are untraceable. When a practitioner completes all the work with PowerPoint and needs to explain it to the team afterwards, it becomes pretty easy for the practitioner to get completely lost in his work. Participant A prefers working with a team using the Powerwall and VR devices rather than working alone using VR devices. While using VR devices is enjoyable, the wearing experience with the headset can be unpleasant due to the discomfort of wearing, motion sickness or other virtual reality sickness.

Regarding future development ideas, participant A suggests that an easy switch in navigation between 2D and 3D modes or between visualisation devices (such as tactile screens, VR headsets, and desktop screens) would be interesting. He perceives that the 3D immersive MBSA interface still requires considerable effort to learn, and existing flaws, such as poor speech recognition functionality, can lead to frustration and impede its usage.

- **Axone**

Both participants in the Axone group participated in stereotyping their software. For the current 2D diagrammatic MBSA interface, they emphasise the complexity of collaborating with other practitioners who use different tools, even for similar activities like functional analysis. In contrast, they identify the possibility of getting different points of view from the same environment as one intriguing advantage of the 3D immersive MBSA interface. Another benefit they noticed with the 3D immersive MBSA interface is that when there are multiple interfaces between two objects, it is possible to differentiate these interfaces by placing them in different locations on the objects with a more specific meaning – e.g., modelling a drug flow that is injected in the abdomen of a 3D patient model. This is related to the additional capacity to convey spatial information with a 3D immersive MBSA interface. They agreed with the

selection of specific models in the model database. The level of realism is well-adjusted, such as the low poly man, which is enough to present one stakeholder in the context without overloading the VR environment. They believe the 3D immersive MBSA interface should not entirely replace the 2D diagrammatic one. Instead, it should be reserved for specific phases of the design process. For instance, when dealing with high-level, abstract elements, the typical scenario involves engineers in different disciplines addressing it with their particular discipline-oriented software for further development. Participant A suggested adding traceability to external entities across different contexts to handle situations with identical external entities in two different contexts.

4.6.7 Formative Usability Testing Discussion

For Schneider Electric, the effort to tailor SysML within Catia Magic at the corporate level was vital as they developed a modelling method and highly customised the ontology and visual vocabulary. Indeed, practitioners in Schneider Electric do not manipulate the SysML concepts (e.g., bloc, property, flow, etc.) but Schneider Electric engineering concepts (services, technical functions, customer requirements, etc.) in a more concise framework. In addition, all concepts are encoded with specific visual metaphors (e.g., shapes, colours, icons, etc.) to increase the readability of models and facilitate learnability and communication, even between non-experts. For the Axone group, it was a similar case with their modelling method. This leads to the result that they have highly rated the 2D diagrammatic MBSA interfaces compared to the Becton Dickinson group.

Moreover, it is worth noting that Participant A is the leader of the development team of this framework and is very satisfied with the adoption of the MBSA framework. Therefore, the 2D diagrammatic MBSA interface satisfies completely participant A, showing from the evaluation results, especially those for research question 1 about User Satisfaction, that he rated the 2D diagrammatic MBSA interface either slightly better or equal to the 3D immersive MBSA interface (with two exceptions: better Visual Expressiveness and Complexity Management on 3D immersive MBSA interface). He mastered the field with his years of experience, and there are barely any barriers to perceptual or conceptual integration with him. His expertise allows him to overcome most of the difficulties we are trying to overcome with the transformation from 2D to 3D. As a result, he did not perceive much difference between 2D and 3D. However, he said the immersion level increased participants' commitment to co-design activities.

Furthermore, all Schneider Electric and Axone participants also scored relatively high for the 3D immersive MBSA interface. Still, only 1 out of 4 (2 out of 6 counting Becton Dickinson) gave a higher score for the 3D immersive MBSA interface than the 2D diagrammatic MBSA

interface, showing that the experts still prefer the 2D diagrammatic MBSA interface as the main approach to work with.

In contrast, one participant from Becton Dickinson poorly rated the 3D immersive MBSA interface on SUS, perceived efficiency, perceived effectiveness, perceived retention rate, intention to use in the future, and task load. As mentioned in the interview, he is very used to Word, Excel, and PowerPoint and is reluctant to use the 3D immersive MBSA interface. VR sickness can be significantly prevented using a 3D immersive MBSA interface.

For visual notation, everyone rated the 3D immersive MBSA interface better than the 2D diagrammatic MBSA interface for every PoN criterion, with two exceptions from the same participant: Axone Participant A for Complexity Management and Graphic Economy. This demonstrates the contribution of a 3D immersive MBSA interface to improving visual notation.

Everyone mentioned replacing the 2D diagrammatic MBSA interface with the 3D immersive MBSA interface, but never in an absolute condition. The consensus is that, with more or less reluctance, they do not want a 3D immersive MBSA interface to replace the 2D diagrammatic MBSA interface entirely. The transformation should be partial and incremental. The 3D immersive MBSA interface should only sometimes be used due to the uncomfortable experience with VR devices and should focus on the preliminary design phase. Additionally, they expect to gain more insights into the 3D immersive MBSA interface, especially once it or an industrial-adapted version is tested with operational pilot projects in the industry.

4.7 Meta-Analysis

A meta-analysis summarises the results across all validation methods for each quality criterion: User Satisfaction, Visual Notation, Presence, and Cognitive Load. Some results are quantitative, whereas others are qualitative, and there is no statistical comparison between these two aspects of the data. The objective is to draw more generic conclusions based on observations from the comparison, indicating a general direction for subsequent research.

4.7.1 User Satisfaction

Table 72 Validation methods and research questions (RQ1)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

The quality criterion of User Satisfaction is measured in single—and multi-person controlled experiments and formative usability tests. The results revealed a difference between the evaluations of the students (novices) and the industrial practitioners (experts).

[SUS] In single-person and multiple-person controlled experiments with the students, the SUS scores for the 3D immersive MBSA interface are significantly higher than the SUS scores for the 2D diagrammatic MBSA interface (see Figure 150). However, in the formative usability tests, only two out of six gave a higher SUS score for the 3D immersive MBSA interface than for the 2D diagrammatic one. Experts also tend to assign higher scores to both interfaces than novices.

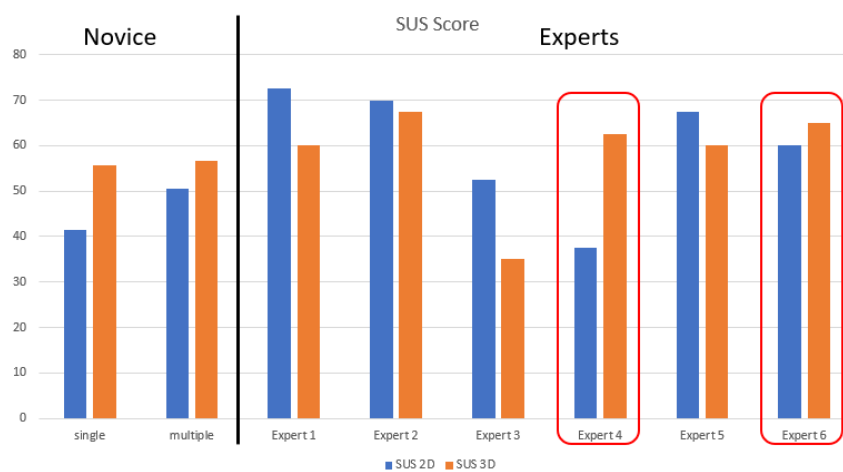


Figure 150 Overview of SUS score

[Perceived Efficiency] In all three validation methods, the perceived efficiency of the 2D diagrammatic MBSA interface is rated either higher or equal to the 3D immersive MBSA interface, except for one expert (expert 6).

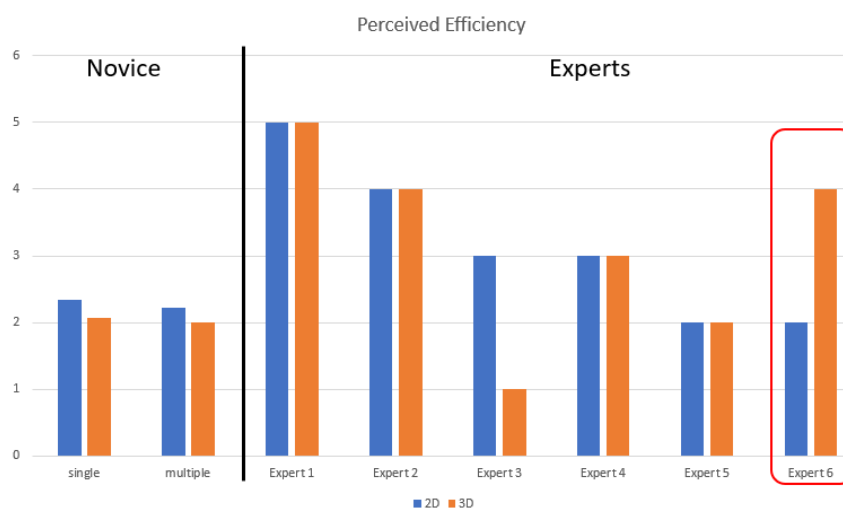


Figure 151 Overview of Perceived Efficiency

[Perceived Effectiveness] Similar to perceived efficiency, the perceived effectiveness of the 2D diagrammatic MBSA interface is rated either higher or equal to the 3D immersive MBSA interface, except for one expert (expert 6).

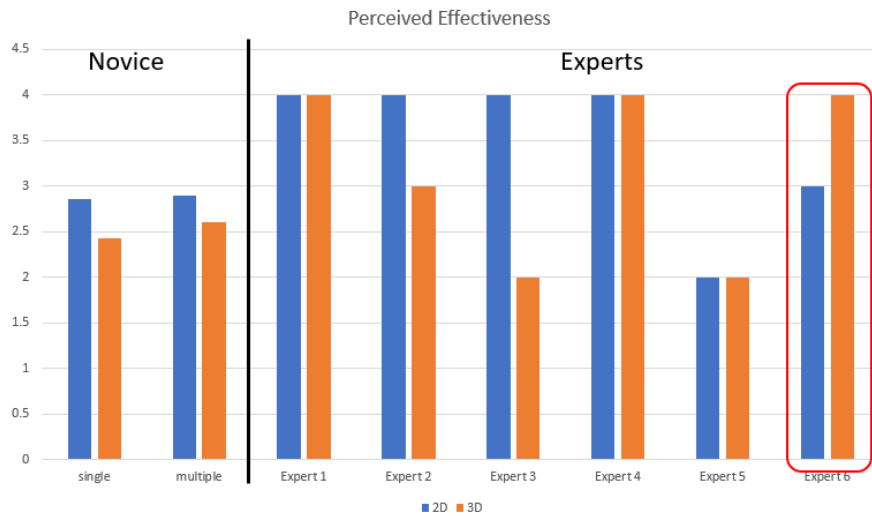


Figure 152 Overview of Perceived Effectiveness

[Perceived Retention Rate] In controlled experiments, novices preferred a 3D immersive MBSA interface over a 2D diagrammatic MBSA interface regarding perceived retention rate. Experts in the formative usability tests showed no preference between a 3D immersive MBSA interface and a 2D diagrammatic MBSA interface for perceived retention rate, except for one expert who preferred a 2D diagrammatic MBSA interface.



Figure 153 Overview of Perceived Retention Rate

[Intention to Use in the Future] Similar to perceived efficiency and perceived effectiveness, users rated higher or equal scores on intention to use in the future for 2D

diagrammatic MBSA interface than 3D immersive MBSA interface, except for one expert (expert 6).

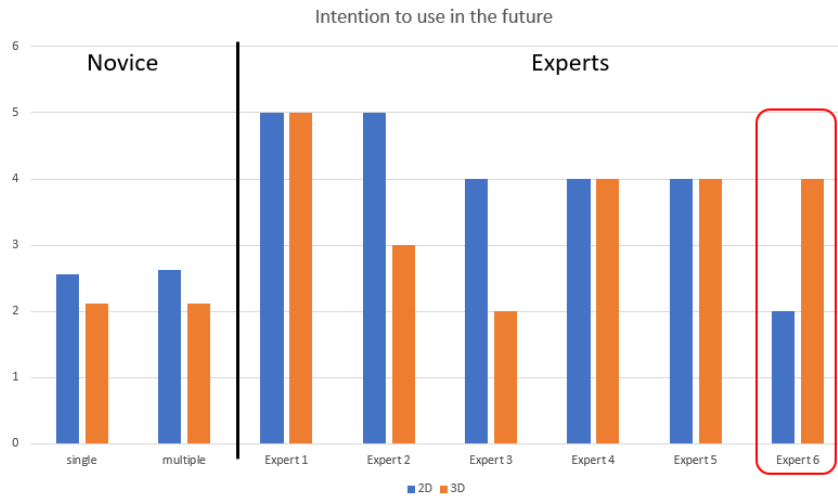


Figure 154 Overview of Intention to use in the future

[Usage Fun] Results in every validation method show a preference for a 3D immersive MBSA interface over a 2D diagrammatic MBSA interface for the Usage Fun.

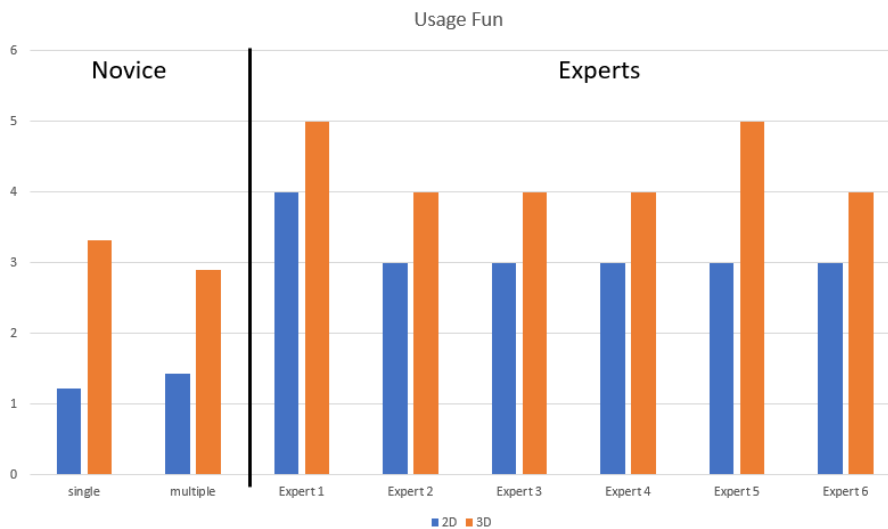


Figure 155 Overview of Usage Fun

[User Motivation] Results in every validation method show a preference for a 3D immersive MBSA interface over a 2D diagrammatic MBSA interface for the User Motivation, except for two experts who do not have a preference.

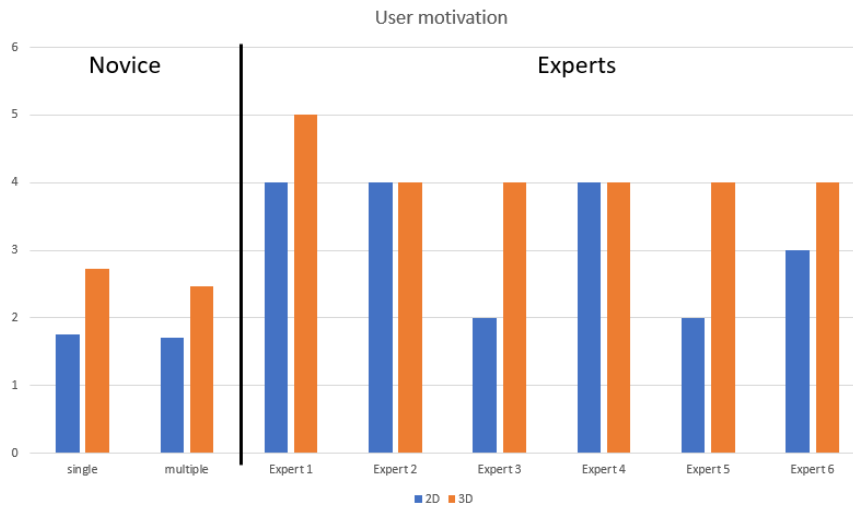


Figure 156 Overview of User Motivation

[Summary] Overall, the 3D immersive MBSA interface has better User satisfaction than the 2D diagrammatic MBSA interface. 3D immersive MBSA interface beats 2D diagrammatic MBSA interface on most criteria except for Perceived Effectiveness, Perceived Efficiency, and Intention to use in the future. The result of an intention to use in the future indicates that even if users prefer the 3D immersive MBSA interface, more is needed to make them shift the modelling interface from a 2D diagrammatic to a 3D immersive MBSA interface. This criterion appears more strongly related to Perceived Efficiency and Perceived Effectiveness. Further research could explore the reason for this and make adjustments to improve Perceived Efficiency and Perceived Effectiveness. From another aspect, the reluctance to adopt the 3D immersive MBSA interface in users' daily work may be due to the unclear declaration. The objective of the 3D immersive MBSA interface is not to entirely replace the 2D diagrammatic but to do so parsimoniously, especially for transdisciplinary co-design activities at a high abstraction level, such as system and system of systems. The results may be skewed towards a 3D immersive MBSA interface with such an indication.

Moreover, two users, Expert 3 and Expert 6, showed outstanding results. Expert 3 showed extremely uncomfortable reactions to VR devices during the experiment. The VR illness broadly interfered with his experience so much that he could not stand the virtual environment for 5 minutes. This is the direct result of his giving the 3D immersive MBSA interface a meagre rating. On the contrary, Expert 6 showed great interest in the 3D immersive MBSA interfaces, rating them higher than the 2D diagrammatic MBSA interfaces even for Perceived Efficiency, Perceived Effectiveness, and Intention to use in the future.

4.7.2 Visual Notation

Table 73 Validation methods and research questions (RQ2)

	<i>RQ1</i> <i>User</i> <i>Satisfaction</i>	<i>RQ2</i> <i>Visual</i> <i>Notation</i>	<i>RQ3</i> <i>Presence</i>	<i>RQ4</i> <i>Cognitive</i> <i>Load</i>	<i>RQ5</i> <i>Group</i> <i>Dynamics</i>
<i>Survey</i>		X			
<i>Expe_S</i>	X	X	X	X	
<i>Expe_M</i>	X	X		X	X
<i>Formative</i>	X	X	X	X	

Every validation method measures the quality criterion Visual Notation. The survey result is discussed separately since it provides limited capture images of different interfaces with different case studies. In contrast, the other validation methods represent two interfaces altogether.

[Perceptual Discriminability] From the survey, we have the result of one case study of a 2D diagrammatic MBSA interface and two different case studies of a 3D immersive MBSA interface (see Figure 157). The 3D immersive MBSA interface can improve Perceptual Discriminability. However, the effect is different in different case studies. Perceptual Discriminability is decreased in the case study of the Fire Forest Detect System. Based on the results of the remaining validation methods (see Figure 158), the 3D immersive MBSA interface has better Perceptual Discriminability compared to the 2D diagrammatic MBSA interface.

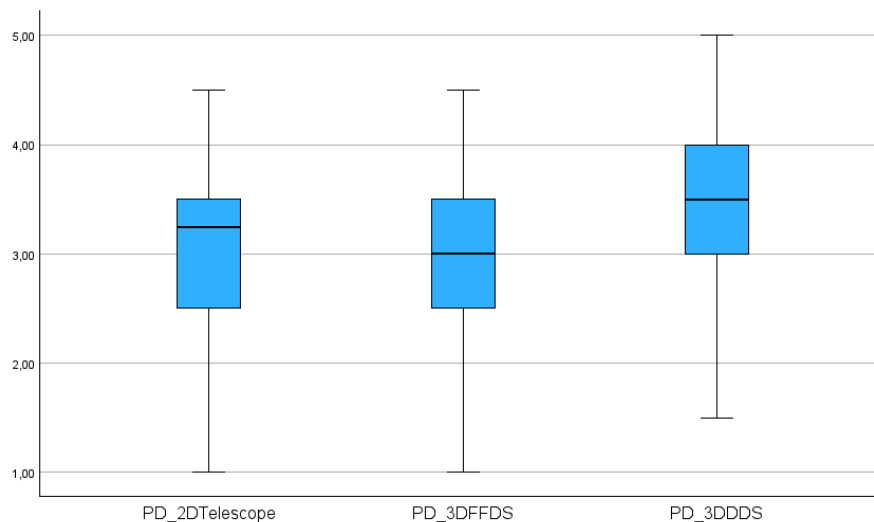


Figure 157 Survey result of Perceptual Discriminability

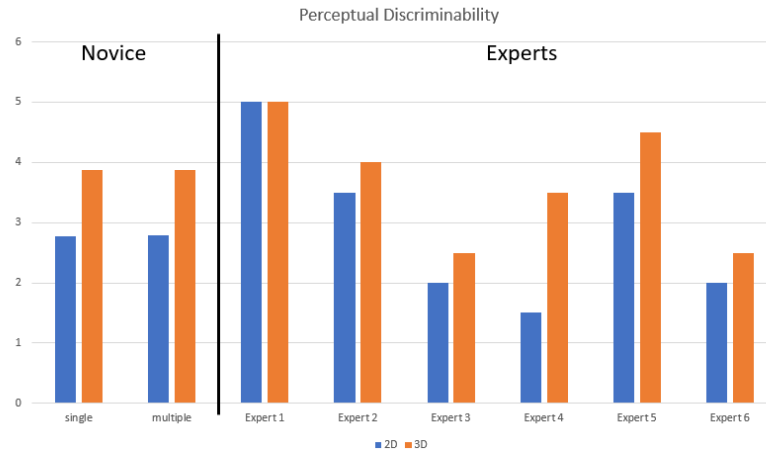


Figure 158 Summary of Perceptual Discriminability

[Semantic Transparency] From the survey, we have the results of two different case studies of a 2D diagrammatic MBSA interface and one case study of a 3D immersive MBSA interface (see Figure 159). The Semantic Transparency of a 2D diagrammatic MBSA interface is similar to the two case studies. A 3D immersive MBSA interface has better Semantic Transparency than a 2D diagrammatic one. Based on the remaining validation methods (see Figure 160), the 3D immersive MBSA interface has better Semantic Transparency than the 2D diagrammatic one.

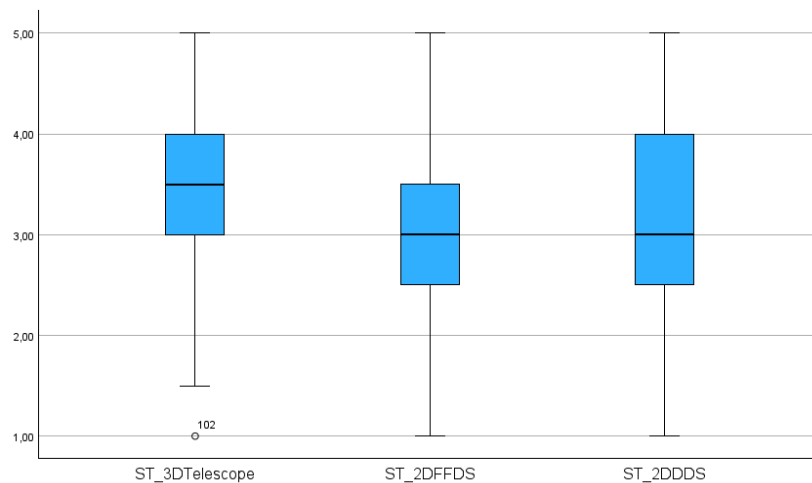


Figure 159 Survey result of Semantic Transparency

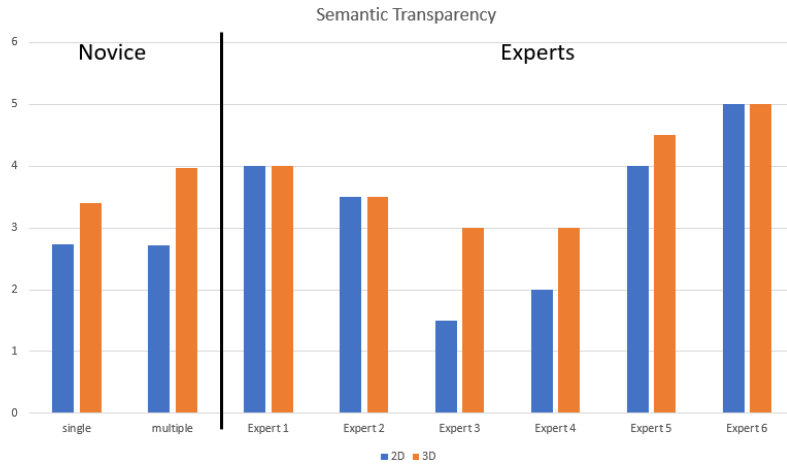


Figure 160 Summary of Semantic Transparency

[Complexity Management] From the survey, we have the results on the 2D diagrammatic MBSA interface (2DFFDS), 2D iconic modelling interface (2DAutomotive), and 3D immersive MBSA interface. (see Figure 161). Complexity Management from a 2D iconic modelling interface and a 3D immersive MBSA interface is similar to and better than a 2D diagrammatic one. Based on the results of the remaining validation methods (see Figure 162), the 3D immersive MBSA interface has better Complexity Management compared to the 2D diagrammatic MBSA interface, with one exception (expert 5).

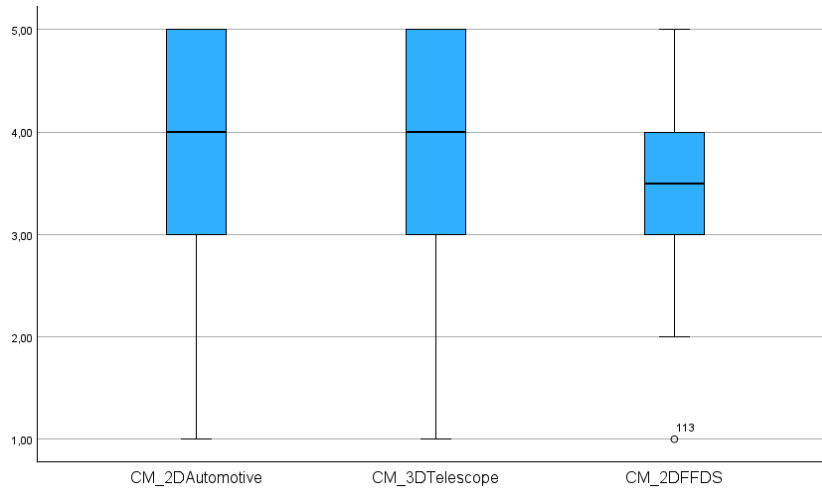


Figure 161 Survey result of Complexity Management



Figure 162 Summary of Complexity Management

[Visual Expressiveness] The survey results include the 2D diagrammatic MBSA interface (2DDDS), 2D iconic modelling interface (2DAutomotive), and 3D immersive MBSA interface (see Figure 163). The 2D iconic and 3D immersive MBSA interfaces are similar to and better than the 2D diagrammatic MBSA interfaces. Based on the remaining validation methods (see Figure 164), a 3D immersive MBSA interface has better Visual Expressiveness than a 2D diagrammatic one.

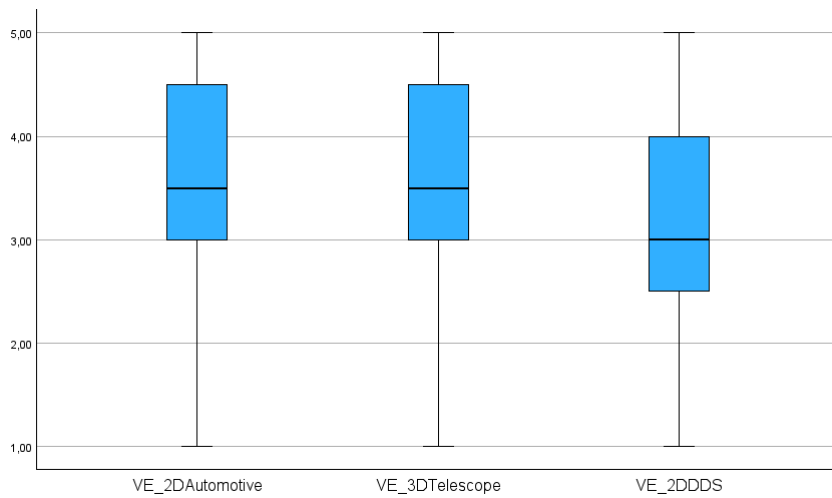


Figure 163 Survey result of Visual Expressiveness

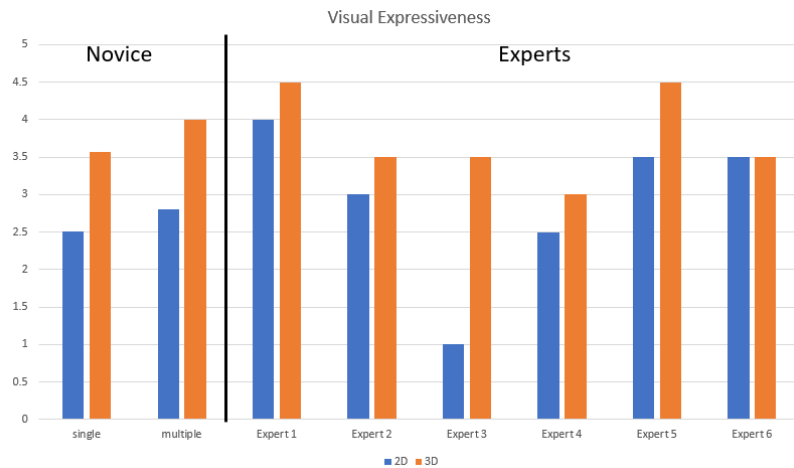


Figure 164 Summary of Visual Expressiveness

[Graphic Economy] From the survey, we have the result of one 2D diagrammatic MBSA interface and two case studies of a 3D immersive MBSA interface (see Figure 165). The graphic Economy from the case study Fire Forest Detect System of 3D immersive MBSA interface is similar to the 2D diagrammatic MBSA interface. Still, the Graphic Economy from the case study Drug Delivery System of 3D immersive MBSA interface is better than the 2D diagrammatic MBSA interface. Based on the results of the remaining validation methods (see Figure 166), the 3D immersive MBSA interface has a better Graphic Economy or equal Graphic Economy compared to the 2D diagrammatic MBSA interface, with two exceptions (expert three and expert 5).

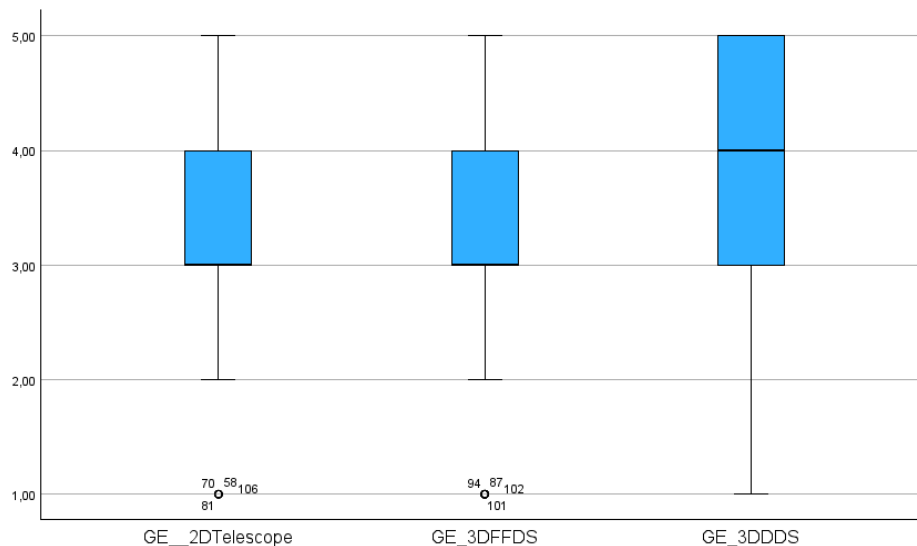


Figure 165 Survey result of Graphic Economy

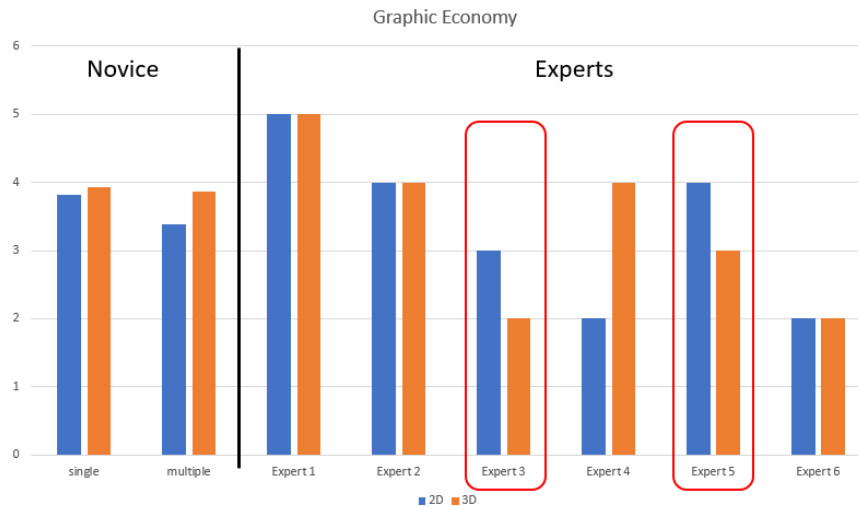


Figure 166 Summary of Graphic Economy

[Summary] Due to the time constraints on the survey online, we are limited in the questions we can include. From what we have explored with the study, the performance of a 2D diagrammatic MBSA interface is more stable than that of a 3D immersive MBSA interface. For specific case studies in 3D immersive MBSA interfaces, some views (i.e., the captured images in the survey) may have difficulties improving certain criteria or, even worse, damaging specific criteria. However, the 3D immersive MBSA interface has a globally better visual notation than the 2D diagrammatic MBSA interface with other validation methods.

4.7.3 Presence

Table 74 Validation methods and research questions (RQ3)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

The quality criterion Presence is measured in Single-person controlled experiments and Formative usability tests.

[Presence] The Presence is in the Not Acceptable range from the single-person controlled experiment, expert 3, and expert 4. Expert 1, expert 2, expert 5, and expert 6 gave an Acceptable Presence score. The black lines in the following four figures (Figure 167, Figure 168, Figure 169, and Figure 170) represent the average scores from the IPQ studies.

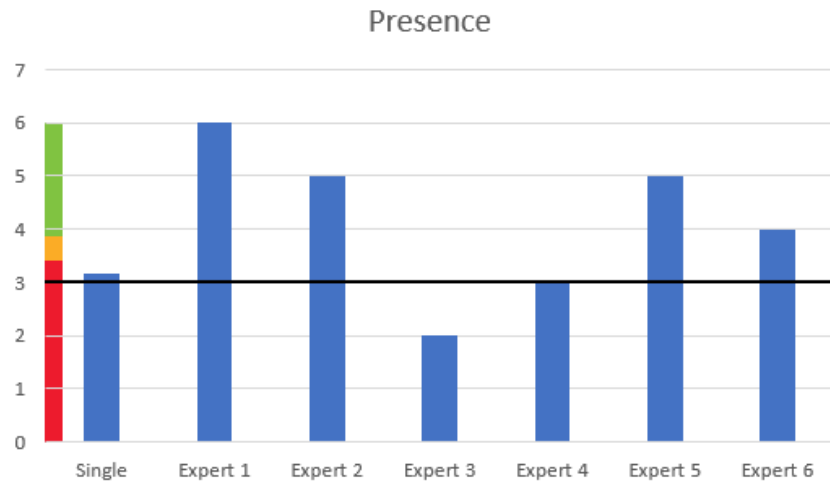


Figure 167 Summary of Presence

[Spatial Presence] The Spatial Presence is in the Not Acceptable range from the single-person controlled experiment, expert 3, and expert 4. Expert 1, expert 2, expert 5, and expert 6 gave an Acceptable Spatial Presence score.

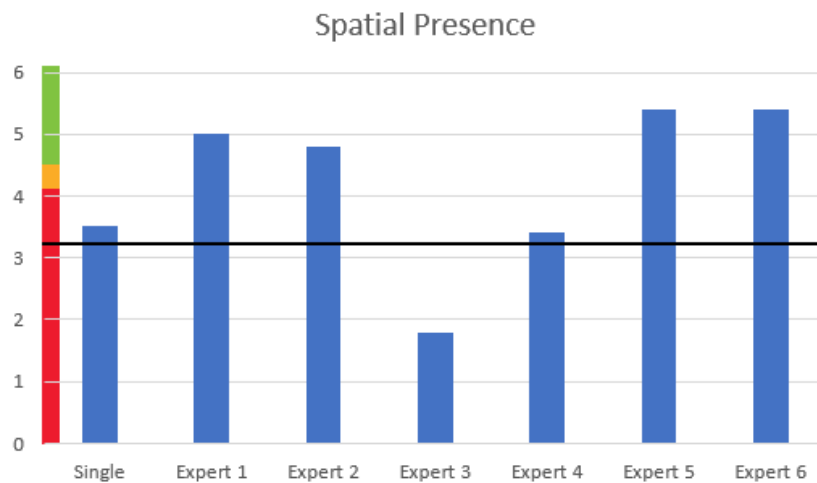


Figure 168 Summary of Spatial Presence

[Involvement] The Involvement is in the Not Acceptable range from the single-person controlled experiment, expert 3, expert 4, and expert 5. Expert 1 and Expert 2 gave an Acceptable Involvement score. Expert 6 gave a Marginally Acceptable Involvement score.

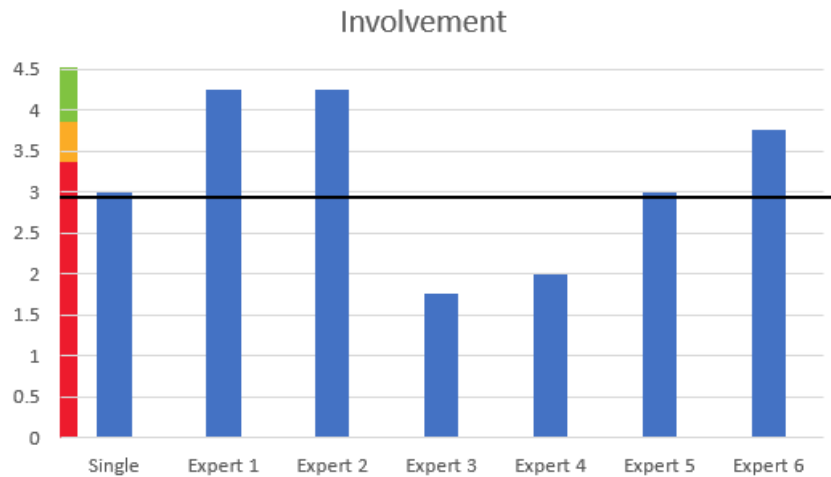


Figure 169 Summary of Involvement

[Experienced Realism] All the results' Experienced Realism scores are in the Not Acceptable range, except for the experts, who rated it as an Acceptable Experienced Realism Score.



Figure 170 Summary of Experienced Realism

[Summary] The average scores of presence criteria from single-person controlled experiments are always in the Not Acceptable range. Expert 1 showed high-level presence engagement by rating all four presence criteria as Acceptable. Expert 3 and Expert 4 consistently rated within the Not Acceptable range. This indicates that the evaluation of Presence is strongly related to individual characteristics. However, the average scores of current studies show that they are consistently below the scores from single-person controlled experiments. This means that the current technologies don't allow us to have an Acceptable or Marginally Acceptable Range with this evaluation approach, or the threshold from this evaluation approach needs to be adopted for the 3D graphics (stereoscopic) via a head-mounted display context.

4.7.4 Cognitive Load

Table 75 Validation methods and research questions (RQ4)

	RQ1 <i>User Satisfaction</i>	RQ2 <i>Visual Notation</i>	RQ3 <i>Presence</i>	RQ4 <i>Cognitive Load</i>	RQ5 <i>Group Dynamics</i>
Survey		X			
Expe_S	X	X	X	X	
Expe_M	X	X		X	X
Formative	X	X	X	X	

The quality criterion Cognitive Load is measured in Single-person controlled experiments, multiple experiments, and Formative usability tests. The Cognitive Load of the 3D immersive MBSA interface is lower than that of the 2D diagrammatic interface, with three exceptions from Expert 1, expert 3, and Expert 5.

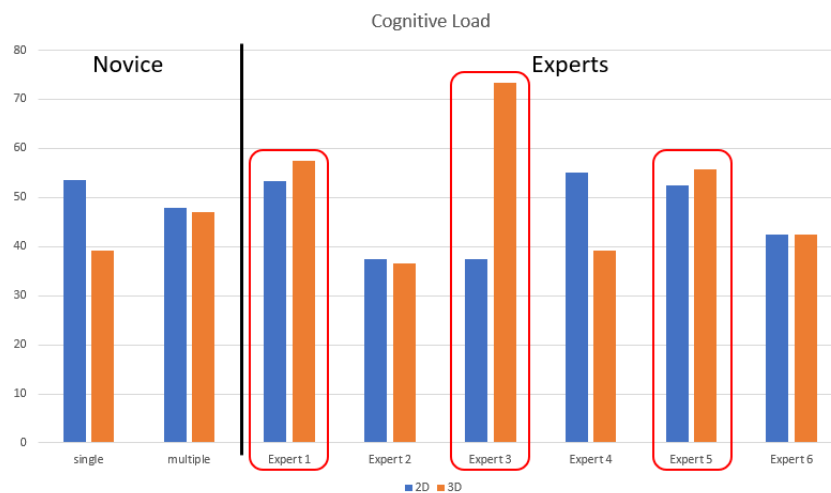


Figure 171 Summary of cognitive load

4.7.5 Summary

The global results of user satisfaction, visual notation, and cognitive load with the students in single-person controlled experiments demonstrate improved quality criteria with the 3D immersive MBSA interface. The difference between the novices and experts is detected in the evaluation of User Satisfaction in the meta-analysis. Still, we can't tell if this difference is significant since it is impossible to conduct a statistical test with a limited number of experts. For other criteria, we failed to detect the difference between novices and experts, including the results from the meta-analysis or the statistical test from the survey. From the literature review, we postulate that there will be a difference between the novices and experts on the cognitive process, which will be reflected in the evaluation results. However, as in the study (Masri, Parker, and Gemino 2008), the Previous Domain Knowledge did not significantly influence the results between novices and experts. VR sickness is another essential factor to consider in

individual cases. For instance, users with a low tolerance to VR sickness, such as expert 3, who rated the 3D immersive MBSA interface with shallow scores, may experience a significant impact on their performance in the VR environment.

5 DISCUSSION

The raw results reported by each validation strategy and the meta-analysis leave no room for the subjective. Thus, before concluding this thesis, I want to discuss the interpretation of the results regarding the proposed research questions, some threats to validity, and some more generic questions.

5.1 Answers to Research Questions

RQ1: Does the 3D immersive MBSA interface increase user satisfaction for co-designing system architecture activities compared with the 2D diagrammatic MBSA interface?

Result: The 3D immersive MBSA interface increases user satisfaction compared to the 2D diagrammatic MBSA interface, but not on every criterion (e.g., perceived effectiveness, perceived efficiency, and intention to use in the future). User satisfaction evaluation is strongly related to VR, which uses users' experiences.

RQ2: Does the 3D immersive MBSA interface provide better visual notations regarding the Physics of Notation (PoN) (Moody 2009) than the 2D diagrammatic MBSA interface?

Result: The 3D immersive MBSA interface has a globally better visual notation than the 2D diagrammatic MBSA interface, but the visual notation evaluation result for the 2D diagrammatic MBSA interface is more stable than that for the 3D immersive MBSA interface across different case studies (i.e., the scores of visual notation criteria for 3D immersive MBSA vary widely across different case studies).

RQ2.1: Does the expertise level in MBSA/SE/SA/VR influence the evaluation of visual notation?

Result: The theoretical answer is yes, but based on the results from the statistical test, it did not demonstrate a significant difference between the experts and novices. It may be due to the evaluation approach (i.e., the grouping of experts and novices or the questions to evaluate their expertise). Future studies can focus on the reason for this observation.

RQ2.2: Does the extra information about the spatial relation provided by the 3D immersive interface improve the MBSA activities?

Result: The spatial relation factor was presented well with a 3D immersive interface to have common user interpretations. However, this conclusion may be anecdotic since only one case study with two representations tested for the spatial relation. For a rounded analysis, a detailed study covering more case studies focusing on spatial relations may be conducted in the future.

RQ3: Does the 3D immersive MBSA interface possess a sufficient level of **presence** so that the users can stay focused on their tasks while in the virtual world?

Result: The 3D immersive MBSA interface provides a poor level of presence (i.e., not acceptable range) based on the scale proposed by Melo. However, the mean scores of the current studies in 3D graphics (stereoscopic) via head-mounted display context are lower than those of the 3D immersive MBSA interface. This means that the current technologies do not enable us to have an Acceptable or Marginally Acceptable Range with this evaluation approach, or the threshold from this evaluation approach needs to be adopted for the 3D graphics (stereoscopic) via a head-mounted display context.

RQ4: Does the 3D immersive MBSA interface reduce the **cognitive load** for co-designing a system architecture compared with a 2D diagrammatic MBSA interface?

Result: The 3D immersive MBSA interface reduces the cognitive load using the NSA-TLX questionnaire compared to the 2D diagrammatic MBSA interface.

RQ5: Does the 3D immersive MBSA interface increase the **group dynamics** for co-designing system architecture activities compared with a 2D diagrammatic MBSA interface?

Result: We failed to answer this question due to students' extremely passive communication behaviour during the experimentation. Future research could focus on preventing this phenomenon by rethinking the experimental conditions.

5.2 Threats to Validity

[Recruitment of experienced participants] The proposed 3D immersive MBSA interface aims to alleviate the current MBSE adoption barriers so people involved in developing future technological systems would be willing to use it frequently. Therefore, the participants should ideally be experienced industrial practitioners for the evaluation. However, recruiting enough system architects to perform statistical tests is almost impossible. Instead of running a controlled experiment with experienced industrial practitioners, we combined data collection and analysis methods with different user profiles. Controlled experiments with students in the university enabled us to conduct a statistical analysis. Formative studies with experienced industrial practitioners followed by a questionnaire and an interview were used to get qualitative feedback.

[Variables in the experiments] As we oriented our research to propose some exciting low-accuracy questions rather than less interesting or uninteresting testable questions, this work assesses various dimensions that influence the potential of using VR for MBSA activities. However, we increased the chance to introduce potential bias during the experimental design. For instance, in the online survey, the independent variable is the type of MBSA interface, and

the dependent variable is the chosen validation criteria. However, variations in the case study (i.e., system of interest) within one type of MBSA interface may influence the evaluation of the criteria, which could lead to potential bias.

[Evaluation through perceived evidence rather than objective factors] In section 4, which details the experiments, due to the time limit and precise identification of means to measure actual properties, most of the dependent variables are the perceived evidence by the participants rather than actual objective factors (e.g., physiological measures) during the experiment. We assume that the participants can correctly perceive information about what we are trying to evaluate throughout the experiments. For instance, the perceived efficiency would correctly reflect the actual efficiency of the MBSA interface. Nevertheless, further investigation must be conducted to confirm or disconfirm whether the perceived evidence correlates to the actual objective evidence.

[Challenge on evaluating group communication performance] Among the criteria of interest to us, we planned to assess the impact of the MBSA interfaces on collaboration, especially communication between participants in the same group. The multi-person experiment gathered students who were reluctant to communicate, making data collection through video and voice recordings impossible.

5.3 Reflections on More Generic Questions

As discussed in the introduction, we had a trade-off between the accuracy of the answers (internal validity) and the breadth of the questions (external validity). In this section, we increase the external validity by reflecting on broader practical questions that motivated this research. The following reflections attempt to provide some elements to the original broad research question: *What is the potential of using VR technology to co-design system architectures, compared with conventional model-based systems architecting notations, software, and devices?*

5.3.1 How do 3D representations differ from diagrammatic ones concerning systems architecting tasks?

The 3D representations in the 3D immersive MBSA interface are more intuitive. They differ from the 2D diagrammatic MBSA interface with the transformation from 2D to 3D, monoscopic to stereoscopic, and symbolic to iconic when adapted. Studies with experiments showing the advantages of that transformation support our hypothesis that a 3D immersive MBSA interface will improve the users' performance. The experiments in this thesis evaluated the diagrammatic representations in a 2D diagrammatic MBSA interface and 3D representations in the 3D MBSA modelling interface with the criteria from PoN. The 3D representations provide

more visual variables (see Figure 172) than 2D diagrammatic representations, such as colour, brightness, shape, texture, and size. Furthermore, apart from the horizontal and vertical positions, it is possible to integrate depth in a stereoscopic 3D space for the planar variables. This increases visual expressiveness and makes it easier for users to discriminate design variables in the environment visually. The constructs also have more intuitive representations with a high level of semantic transparency. In other words, the transformation from symbolic to iconic representations leads to a more straightforward interpretation of the meaning from its appearance.

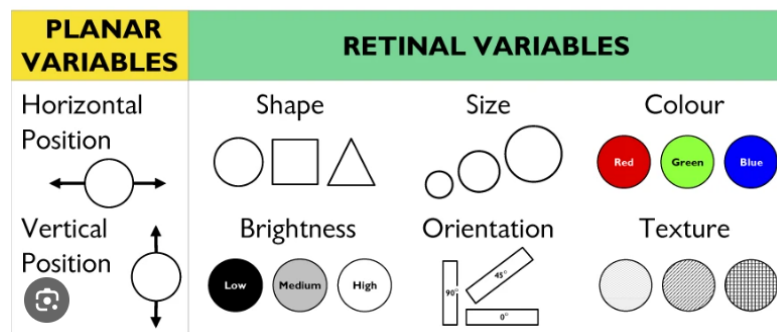


Figure 172 Bertin's eight visual variables (D. Moody 2009)

5.3.2 How did the immersive MBSA interface help architects complete their tasks?

From one aspect discussed in 5.3.1, the 3D representations help users ease perceptual progress by reducing the extraneous cognitive load (see Figure 173). This allows users to capitalise on the conceptual process with limited working memory capacity. Furthermore, apart from helping the perceptual progress, the 3D immersive MBSA interface also allows the conceptual process. The possibility of having all the elements in one single environment enforces the interface's consistency. The shift of views during navigation will not interrupt the users' cognitive process as in the 2D diagrammatic MBSA interface. It is also much easier for a user to understand how one element relates to the whole system since all elements exist within a unified environment, so the transitioning views do not break the consistency. In other words, a view at the systemic level $n+1$ is a long shot for an individual element at the systemic level n . Moreover, limited texts indicate the systemic level and semantic construct (e.g., lifecycle, context, system, etc.). The objective is to provide the system architect with a "map" (as in the general sense of a map in the real world or video games), allowing them to quickly locate themselves in the project. Further research on a more efficient "map" will significantly enhance the conceptual cognitive process. The immersive dimension enables the users to immerse themselves in the virtual environment to help them concentrate on their tasks, assuming they do not experience excessive VR sickness.

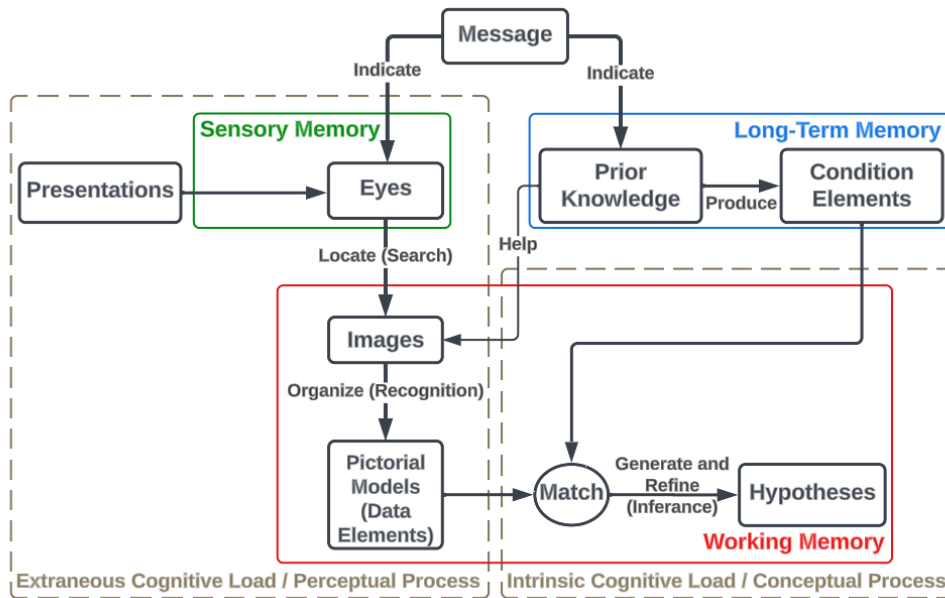


Figure 173 Illustration of diagrammatic reasoning with visual system

A question evaluates the retention rate. Results from several validation methods show an advantage with the perceived retention rate (see Figure 174). The single-person and multiple-person controlled experiments provide statistical evidence. However, all experts rated equality between 2D diagrammatic and 3D immersive MBSA interfaces, with one exception rated worse for 3D immersive. It may reflect the “true” retention rate, or in another case, it may be due to the interference of their previous domain knowledge on the perceived result of the retention rate. There was no specific evaluation of the actual retention rate with objective facts, but we noticed something while modelling the Forest Fire Detection system (FFDS) context. There are around three groups containing 15 elements. While building the diagram in 2D, we occasionally checked the case study even though we had already built the same diagram. However, after completing the context in 3D, we had no problem making the entire context with the correct grouping precision after the initial construction. This supports the hypothesis that, in our case, the 3D immersive MBSA interface can help memorise the content. Future research to evaluate the retention rate with objective facts to generalise this observation would hopefully prove the advantage of a 3D immersive MBSA interface.



Figure 174 Overview of Perceived Retention Rate

5.3.3 Is there a difference between notational experts and novices?

Theoretically, the prior knowledge will affect the cognitive process, resulting in differences between experts and novices. The difference here refers to how experts and novices evaluate the modelling interface and how they evaluate the difference between a 2D diagrammatic MBSA interface and a 3D immersive MBSA interface. As illustrated in the reference source, it was not found. The experts rely on their prior knowledge of the perceptual and conceptual processes. The intrinsic and extraneous cognitive load are often discussed when referring to cognitive load. The third type, germane cognitive load, was later proposed by Sweller (Sweller, Van Merriënboer, and Paas 1998). Germane cognitive load refers to the mental effort required to process and integrate new information into existing knowledge structures, leading to meaningful learning and long-term retention. Prior knowledge can influence it. Those theories (Sweller and Chandler 1994; Mayer 2001; Khatri et al. 2006) support the hypothesis that there is a difference between the experts and novices, but the statistical test results in this thesis failed to corroborate the theory. As discussed in (Masri, Parker, and Gemino 2008), it may be due to the instrument used to measure prior knowledge and, in this study, it may also be due to the limited number of expert participants.

5.3.4 What barriers did users face when using the immersive MBSA interface?

One significant barrier would be VR sickness, also known as cybersickness or simulator sickness, referring to the discomfort or nausea when using VR devices. The representative case is the third expert in the formative usability tests experiment. If the user has a low tolerance for VR sickness, it becomes a dominant factor hindering them from completing their tasks. This can be alleviated with the development of VR technology (Huygelier et al. 2019).

For instance, it is possible to reduce VR sickness by improving the display quality, frame rates, and motion tracking accuracy. In a 3D immersive MBSA interface, changing the navigation mode (walk, fly, and teleport) and adjusting the movement speed is possible. Furthermore, taking a break from VR devices can prevent potential symptoms of VR sickness, so we introduced the 2D mode with a keyboard and mouse for specifying the requirements. We would not expect people to spend more than 45 minutes in the headset. Still, we propose that users gather for short co-design sessions before continuing the tasks in detail using a desktop-based working environment. Finally, it is more difficult for users who are new to VR to adapt to a VR environment. They need to spend more time and effort getting familiar with the interaction. Some designs, such as embedded tutorials and tips for buttons to interact with, could assist users in quickly familiarising themselves with the 3D immersive MBSA interface.

5.3.5 Did the immersive MBSA interface facilitate co-design activities?

The feedback on the collaborative performance was mainly from the experts, as it was challenging with the multiple-person controlled experiment. Experts claim to have a more engaging and interactive design experience with the Powerwall, and it facilitated a shared understanding among the participants in the collaborative systems architecting process. The immersive aspect gives them a more thorough system perspective, positively impacting the architecture design activity. The experts in front of the Powerwall gave feedback and modelling instructions to the system modeller immersed in the VR headset. They did not have to put effort into the modelling tasks. Thus, it seems that when core team members are co-designing a system architecture, having someone in charge of the modelling tasks can enable them to reallocate the saved cognitive load to fundamental systems architecting tasks.

5.3.6 Could people imagine using a similar immersive MBSA interface?

Generally, the answer to this question is yes, with some conditions. Students and experts highly rated the 3D immersive MBSA interface in terms of user satisfaction and visual notation. Most of them think the 3D immersive MBSA interface reduces cognitive load. We should have evaluated group dynamics, but students who participated in the multi-user experiment were excited to experience a virtual environment. They thought the 3D immersive MBSA interface was intuitive, simple, and fun. A few of them asked if it is possible to integrate this interface into the teaching session, which can be a research perspective in the future. Indeed, it is very challenging to teach abstract MBSA concepts and methodological practices to students without any experience in new product development and industry in general, and 3D is a strong candidate to make those concepts more tangible.

Experts who have spent many years deploying MBSE in their organisation prefer a smooth transition, not a raw and complete transformation from a 2D diagrammatic to a 3D immersive

MBSA interface. In other words, they prefer an incremental evolution rather than an instant and total revolution. They also imagine themselves using both interfaces alternately. For instance, they believe that the 3D immersive environment with a core team standing in front of the Powerwall is interesting for conducting co-design sessions where elements resulting from different viewpoints are efficiently gathered in a unique space. Indeed, today, the scope of MBSA is often limited to a few R&D actors, whereas systems architecting requires the integration of multiple perspectives. We did not have the time to conduct experiments with a core team, including representatives from R&D, marketing, quality, operations, project management, and top management. Still, the results motivated the hypothesis that the 3D immersive MBSA would facilitate systems architecting tasks compared to diagrammatic representations. Moreover, experts suggested using the 3D immersive MBSA interface to make the working atmosphere more ludic and lighter. Although it may be assumed as a secondary result, increasing the willingness to use MBSA tools is essential, knowing that existing conventional solutions require architects to spend more time creating and polishing diagrams than engineering systems. Finally, the system architects who contributed to this study reported that the 3D immersive MBSA interface is of great interest for system-level analyses, including the system decomposition into subsystems. When defining detailed and concrete subsystems or components, detailed design often comes with dedicated engineering software. Of course, this observation depends on how abstract the system of interest is. Architecting a future air combat system is not architecting a drug delivery system. In the future, the 3D immersive interface could be extended to develop a solution that supports activities for architecting systems of systems, such as co-architecting an aircraft and its industrial system.

5.3.7 How would user improve the MBSA interface to integrate it into their work?

Based on the results of all evaluation methods, here are a few points for the improvement of the 3D immersive MBSA interface:

- **Representation for function:** Users were confused about the functions because they did not visually see them after their creation. The functions are presented as verbal information in the VR environment. Users can play the oral functions by pressing the menu's play button, but no visual representations exist for the created functions. They occasionally get lost because of this. Adding a 3D representation to present the created function would help in the future. Moreover, a list of functions with hierarchical sub-functions would enhance the management and exploration of existing functions.
- **Adaptation to field activities:** Regarding the property of MBSA as an interdisciplinary approach, it has a wide field of practice. To maintain the 3D immersive MBSA interface efficiently and effectively, some adaptations related to the practice field would be an

improvement. This echoes the idea in PoN as cognitive fit, which is defined as different representations of information suitable for other tasks and audiences. Such adaptations dedicated to the tasks and users can be applied to the representations (i.e., visual graphics of the modelling language) and the semantic constructs. Moreover, the scope can extend to the framework and software as well.

- **Better navigation:** One advantage of the 3D immersive MBSA interface is its ability to represent all information within one VR environment, which enhances project navigation. Experts thought there was room for improvement in navigating the VR environment after using the 3D immersive MBSA interface. The navigation transition could be smoother with the help of 3D space, and the navigation mode can be more powerful, e.g., allowing users to glimpse other contexts while remaining in the current one.
- **Traceability of elements:** In the literature review on diagrammatic reasoning, we discussed the cognitive process as perceptual and conceptual progress. As discussed in previous question 5.3.2, perceptual progress is one significant contribution of the current 3D immersive MBSA interface. Showing how model elements are connected in different contexts and lifecycle phases will also help the conceptual progress.
- **Adding reusability:** Experts suggested implementing functionality to reuse existing built elements. For instance, a system with sub-systems designed could be reused in another context, or elements imported in one context can be reviewed and dragged into a new context. Version control with capabilities to manage options, variants, and design alternatives could enhance this functionality.
- **More efficient interaction:** Users noticed that specific interactions in the prototype were somewhat impractical. For instance, despite VR easing the manipulation of 3D representations in the VR environment (e.g., the user can easily move the 3D representations to the desired position in the 3D space compared to the conventional way with the PC screen), users spend time aligning all the elements in a VR environment. Moreover, the current interaction for rotation does not efficiently allow users to rotate 3D representations (e.g., the user wants to rotate the 3D representation of one stakeholder on the y-axis) from a distance. Further studies on efficient interaction metaphors could improve the 3D immersive MBSA interface.

6 CONCLUSION

This thesis explores the potential benefits of VR technology in the context of MBSA with a new 3D immersive Model-Based Systems Architecting (MBSA) interface for co-designing systems architecture. The proposed 3D immersive MBSA interface offers an intuitive modelling language, a comprehensive modelling approach, and an immersive Virtual Reality (VR) environment. The modelling interface decreases learning time and effort with simple, efficient, intuitive, and ludic interactions. The methodology is MBSA-based, with the common denominator of existing frameworks, new 3D representations of the semantic constructs, and a VR environment with implemented interactions.

6.1 Contribution and Claim

The proposed 3D immersive MBSA interface combines the MBSA approach and VR technology. The combination is not a transition of 2D diagrams into 3D VR space. The 3D immersive MBSA interface harnesses the advantages of the VR environment. It has a new visual syntax, incorporating common semantic constructs derived from popular MBSA methodologies. The 3D representations are more cognitively efficient than the conventional 2D diagrams. The new tool that stages the 3D representations as an immersive environment and interface is easy to learn, with intuitive interactions. Several criteria serve to evaluate the 3D immersive environment across different evaluation strategies. The global result is a promising start for further research to understand how humans represent and interact with different types of MBSA information and to develop new human-computer interactions to advance human-centred computing in MBSA.

On the other hand, this thesis explores the interest in integrating VR into MBSA, such as the proposed 3D immersive MBSA interface. We compared it with 2D diagrammatic MBSA interfaces to evaluate its performance. However, it does not mean we want to completely replace the traditional MBSA diagrams with 3D immersive human-computer interaction. The 3D immersive MBSA interface could focus on the preliminary activities of the architecture definition process where all stakeholders' voices are necessary. It could be an alternative modelling modality for a change to a 2D diagrammatic MBSA interface (and vice versa) or even a ludic learning approach for the MBSA novices.

The evaluation part of this thesis was challenging, as it is also one aspect that has always been ignored in immersive HCI for MBSA. The scientific support for the experiments could be more solid due to the limited references and research time. However, we do not aim to strive for flawless experiments with sound scientific processes. As exploratory research, we prefer

to work on the experiment from a general idea to demonstrate the interest of VR technology in the collaborative architecture design of engineered systems.

6.2 Perspectives: A Vision of VR for MBSA

VR technology is drawing more and more attention, and its development has exploded in recent years. However, VR users still experience issues, primarily from VR sickness. Along with the development of VR technology, VR sickness problems will be alleviated or even eliminated. It will boost all applications within VR fields and enable co-working from various devices and locations when teleworking becomes a de-facto standard. MBSA practitioners will have a much better user experience, and a 3D immersive MBSA interface tightly connected to the digital ecosystem will be the new conventional way in the community.

Furthermore, SysML V2 was recently released. The proposed 3D immersive MBSA interface is also compatible with SysML V2, and their ideas for improvement or motivation for improvement are similar, too, such as enhancing consistency and traceability and making the language more intuitive and novice-friendly. SysML V2 is no longer based on UML but on a new foundation technology called kernel-modelling language. It provides not only a base of specific language constructs but also the semantics of these constructs, enabling system engineering capabilities to an extensible approach, bimodal language as textual and graphic language, expressive and precise decision-making, and core patterns for novice-friendly language versus V1. The built-in API of SysML V2 allows it to connect easily to other tools, which will promote the market with many more tool vendors. This revolution will boost the studies on the MBSA and empower system engineers to develop advanced technology based on a wide range of disciplines.

Based on the discovery of this thesis, here are several perspectives for the future research:

- Conduct the controlled experiment with fewer independent variables for more solid results for the proposed research questions proposed in this thesis (e.g., one type of 2D diagram in 2D diagrammatic MBSA interface and the informational equivalence representation in 3D immersive MBSA interface for the same system of interest). For instance:
 - RQ:** Does the operational scenario for a drug delivery system in a 3D immersive MBSA interface provide better visual notations regarding the PoN than the 2D diagrammatic MBSA interface?
- Find a more effective approach to evaluate the group communication performance:

RQ: What is the most efficient and effective approach to assessing group communication performance for 3D immersive and 2D diagrammatic MBSA interfaces?

- Explore and confirm the relation between the perceived and actual performance of the criteria:

RQ: What is the relation between the perceived and actual performance of the criteria used to evaluate the 3D immersive and 2D diagrammatic MBSA interface?

- Investigate the impact of the 3D immersive MBSA interface on both concrete and abstract systems and detect the difference between them:

RQ: How does the 3D immersive MBSA interface impact the MBSA activities for concrete and abstract systems, and what is the difference?

- Conduct follow-up research to explore the reasons behind the lower perceived efficiency and effectiveness in the 3D immersive MBSA interface:

RQ: Why do users perceive lower efficiency and efficacy for a 3D immersive MBSA interface than a 2D diagrammatic MBSA interface?

Finally, if we dare to be ambitious, the future of MBSA involves systems architects orchestrating human-centred design with generative AI that supports the authoring and management of MBSA data. Virtual worlds are interconnected in a Metaverse that operates across various mixed reality environments, resembling the high-tech designing tools depicted in science fiction movies.

REFERENCES

- Albers, Albert, and Christian Zingel. 2013. "Challenges of Model-Based Systems Engineering: A Study towards Unified Term Understanding and the State of Usage of SysML." In *Smart Product Engineering*, edited by Michael Abramovici and Rainer Stark, 83–92. Lecture Notes in Production Engineering. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-30817-8_9.
- Andersson, Henric, Erik Herzog, Gert Johansson, and Olof Johansson. 2009. "Experience from Introducing Unified Modeling Language/Systems Modeling Language at Saab Aerosystems." *Systems Engineering* 13 (November): 369–80. <https://doi.org/10.1002/sys.20156>.
- Anwer, Sajid, and Mohamed El-Attar. 2014. "An Evaluation of the Statechart Diagrams Visual Syntax." *International Conference on Information Science and Applications*, 1–4. <https://doi.org/10.1109/ICISA.2014.6847358>.
- Averbukh, V. L., A. U. Baydalín, D. R. Ismagilov, A. U. Kazantsev, and S. P. Timoshpolskiy. 2004. "Utilizing 3d Visualization Metaphors." In *Proceedings of THE 14th International Conference on Computer Graphics & Vision GRAPHICON*, 295–98.
- Averbukh, Vladimir, Natalya Averbukh, Pavel Vasev, Ilya Gvozdarev, Georgy Levchuk, Leonid Melkozerov, and Igor Mikhaylov. 2019. "Metaphors for Software Visualization Systems Based on Virtual Reality." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11613 LNCS: 60–70. https://doi.org/10.1007/978-3-030-25965-5_6.
- Bach, C, and D L Scapin. 2003. "Adaptation of Ergonomic Criteria to Human-Virtual Environments Interactions."
- Baduel, Ronan, Mohammad Chami, Jean Michel Bruel, and Iulian Ober. 2018. "SysML Models Verification and Validation in an Industrial Context: Challenges and Experimentation." *The European Conference on Modelling Foundations and Applications, ECMFA 10890 LNCS*: 132–46. https://doi.org/10.1007/978-3-319-92997-2_9.
- Bangor, Aaron, Philip T. Kortum, and James T. Miller. 2008. "An Empirical Evaluation of the System Usability Scale." *International Journal of Human-Computer Interaction* 24 (6): 574–94. <https://doi.org/10.1080/10447310802205776>.
- Bangor, Aaron, Technical Staff, Philip Kortum, James Miller, and Technical Staff. 2009. "Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale." *Journal of Usability Studies* 4 (3): 114–23.
- Barosan, Ion, and Juan Van Der Heijden. 2022. "Integration of SysML Models in a 3D Environment for Virtual Testing and Validation." In *Federated Africa and Middle East Conference on Software Engineering*, 39–45. Cairo-Kampala Egypt: ACM. <https://doi.org/10.1145/3531056.3542784>.
- Barth, A., Emmanuel Caillaud, and Bertrand Rose. 2011. "How to Validate Research in Engineering Design?" *International Conference on Engineering Design ICED11, Denmark 2* (January): 1–11.
- Bertin, Jacques. 1983. *Semiology of Graphics*. 1er édition. Redlands, Calif: ESRI Press.
- Biggs, Geoffrey, Tomas Juknevičius, Andrius Armonas, and Kyle Post. 2018. "Integrating Safety and Reliability Analysis into MBSE: Overview of the New Proposed OMG Standard." *INCOSE International Symposium* 28 (1): 1322–36. <https://doi.org/10.1002/j.2334-5837.2018.00551.x>.
- Blackwell, A. F., C. Britton, A. Cox, T. R.G. Green, C. Gurr, G. Kadoda, M. S. Kutar, et al. 2001. "Cognitive Dimensions of Notations: Design Tools for Cognitive Technology." *Lecture*

- Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)* 2117 (January): 325–41. https://doi.org/10.1007/3-540-44617-6_31.
- Bolles, Edmund Blair. 1991. *A Second Way of Knowing: The Riddle of Human Perception*. Prentice Hall Press.
- Bowman, Doug A., Donald B. Johnson, and Larry F. Hodges. 2001. "Testbed Evaluation of Virtual Environment Interaction Techniques." *Presence: Teleoperators and Virtual Environments* 10 (1): 75–95. <https://doi.org/10.1162/105474601750182333>.
- Bowman, Doug A., David Koller, and Larry F. Hodges. 1997. "Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques." *IEEE Annual International Symposium Virtual Reality* 46 (9): 1486–98. <https://doi.org/10.1108/K-11-2016-0308>.
- Boy, Guy André. 2017. "Human-Centered Design of Complex Systems: An Experience-Based Approach." *Design Science* 3: e8. <https://doi.org/10.1017/dsj.2017.8>.
- Brooke, John. 1996. "SUS - A Quick and Dirty Usability Scale."
- Brooks, F.P. 1999. "What's Real about Virtual Reality?" *IEEE Computer Graphics and Applications* 19 (6): 16–27. <https://doi.org/10.1109/38.799723>.
- Brooks, Frederick P. 1988. "Grasping Reality through Illusion—Interactive Graphics Serving Science." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1–11.
- Campo, Kelly X., Thomas Teper, Casey E. Eaton, Anna M. Shipman, Garima Bhatia, and Bryan Mesmer. 2023. "Model-based Systems Engineering: Evaluating Perceived Value, Metrics, and Evidence through Literature." *Systems Engineering* 26 (1): 104–29. <https://doi-org.sid2nomade-2.grenet.fr/10.1002/sys.21644>.
- Card, Stuart K., ed. 1983. *The Psychology of Human-Computer Interaction*. Boca Raton: CRC Press. <https://doi.org/10.1201/9780203736166>.
- Carmines, Edward G., and Richard A. Zeller. 1979. *Reliability and Validity Assessment*. Sage publications.
- Chami, Mohammad, and Jean-Michel Bruel. 2018. "A Survey on MBSE Adoption Challenges." *INCOSE Systems Engineering Conference of the Europe, Middle-East and Africa Sector EMEASEC 2018* (November).
- Chandler, Daniel. 2007. *Semiotics: The Basics*. 2nd ed. Basics (Routledge (Firm)). London ; New York: Routledge.
- Chandrasekaran, B., Janice Glasgow, and N Hari Narayanan. 1995. "Diagrammatic Reasoning." MIT Press. 1995. <https://mitpress.mit.edu/9780262571128/diagrammatic-reasoning/>.
- Cloutier, Robert. 2015. "Current Modeling Trends in Systems Engineering." *INSIGHT* 18 (August): 10–13. <https://doi.org/10.1002/inst.12013>.
- Darses, Françoise, Françoise Détienne, Pierre Falzon, and Willemien Visser. 2001. "A Method for Analyzing Collective Design Processes." In *ECCE*, 10:47–56.
- Darses, Françoise, and Pierre Falzon. 1996. "La Conception Collective: Une Approche de l'ergonomie Cognitive."
- Delligatti, Lenny. 2013. *SysML Distilled: A Brief Guide to the Systems Modeling Language*. 1st ed. Addison-Wesley Professional.
- Di Gironimo, G., G. Matrone, A. Tarallo, M. Trotta, and A. Lanzotti. 2013. "A Virtual Reality Approach for Usability Assessment: Case Study on a Wheelchair-Mounted Robot

- Manipulator.” *Engineering with Computers* 29 (3): 359–73. <https://doi.org/10.1007/s00366-012-0274-x>.
- Dionisio, John David N., William G. Burns Iii, and Richard Gilbert. 2013. “3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities.” *ACM Computing Surveys (CSUR)* 45 (3): 1–38.
- Dix, Alan, ed. 2003. *Human-Computer Interaction*. 3. ed., [6. Nachdr.]. Harlow: Pearson Prentice-Hall.
- Dori, Dov. 2002. *Object-Process Methodology – A Holistic Systems Paradigm*. New York. <https://doi.org/10.1007/978-3-642-56209-9>.
- . 2011. “Object-Process Methodology for Structure-Behavior Co-Design.” In *Handbook of Conceptual Modeling*, edited by David W. Embley and Bernhard Thalheim, 209–58. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15865-0_7.
- Dori, Dov, Iris Reinhartz-Berger, and Arnon Sturm. 2003. “Developing Complex Systems with Object-Process Methodology Using OPCAT.” In *Conceptual Modeling - ER 2003*, edited by Il-Yeol Song, Stephen W. Liddle, Tok-Wang Ling, and Peter Scheuermann, 2813:570–72. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-39648-2_46.
- Dumas, Angela, and Henry Mintzberg. 1991. “Managing the Form, Function, and Fit of DESIGN.” *Design Management Journal (Former Series)* 2 (3): 26–31. <https://doi.org/10.1111/j.1948-7169.1991.tb00573.x>.
- Eco, Umberto. 1976. *A Theory of Semiotics*. Indiana University Press. <https://www.jstor.org/stable/j.ctt16xwcf4>.
- El-Attar, Mohamed. 2019. “Evaluating and Empirically Improving the Visual Syntax of Use Case Diagrams.” *Journal of Systems and Software* 156: 136–63. <https://doi.org/10.1016/j.jss.2019.06.096>.
- Emes, M. R., P. A. Bryant, M. K. Wilkinson, P. King, A. M. James, and S. Arnold. 2012. “Interpreting ‘Systems Architecting.’” *Systems Engineering* 15 (4): 369–95. <https://doi.org/10.1002/sys.21202>.
- Engineering, International. 2015. *INCOSE Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*.
- Estefan, Jeff A. 2007. “Survey of Model-Based Systems Engineering (MBSE) Methodologies.” *IncoSE MBSE Focus Group* 25 (8): 1–12.
- Fernandez, Jose L., and Carlos Hernandez. 2019. *Practical Model-Based Systems Engineering*.
- Figl, Kathrin, and Michael Derntl. 2011. “The Impact of Perceived Cognitive Effectiveness on Perceived Usefulness of Visual Conceptual Modeling Languages.” In *Conceptual Modeling – ER 2011*, edited by Manfred Jeusfeld, Lois Delcambre, and Tok-Wang Ling, 6998:78–91. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24606-7_7.
- Foley, James D. 1996. “12.7 Constructive Solid Geometry.” *Computer Graphics: Principles and Practice*, 533–58.
- Formentini, Giovanni, Francois Bouissiere, Claude Cuiller, Pierre Eric Dereux, and Claudio Favi. 2022. “CDFA Method: A Way to Assess Assembly and Installation Performance of Aircraft System Architectures at the Conceptual Design.” *Research in Engineering Design* 33 (1): 31–52. <https://doi.org/10.1007/s00163-021-00378-5>.

- Forsberg, Kevin, and Harold Mooz. 1991. "The Relationship of System Engineering to the Project Cycle." In *INCOSE International Symposium*, 1:57–65. Wiley Online Library.
- Friedenthal, Sanford, Regina Griego, and Mark Sampson. 2009. "INCOSE Model Based Systems Engineering (MBSE) Initiative."
- Friedenthal, Sanford, Alan Moore, and Rick Steiner. 2014. *A Practical Guide to SysML: The Systems Modeling Language*. Morgan Kaufmann.
- Gabbard, J.L., D. Hix, and J.E. Swan. 1999. "User-Centered Design and Evaluation of Virtual Environments." *IEEE Computer Graphics and Applications* 19 (6): 51–59. <https://doi.org/10.1109/38.799740>.
- Gero, John S., and Udo Kannengiesser. 2004. "The Situated Function-Behaviour-Structure Framework." *Design Studies* 25 (4): 373–91. <https://doi.org/10.1016/j.destud.2003.10.010>.
- Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols*. Hackett publishing.
- Grundel, Martin, and Jutta Abulawi. 2016. "SkiPo – a Sketch and Flow Based Model to Develop Mechanical Systems." *INCOSE International Symposium* 26 (1): 399–414. <https://doi.org/10.1002/j.2334-5837.2016.00168.x>.
- Guillermo, Hugo, and Chalé Góngora. 2014. "MBSE for Railway Rolling Stock - Alstom ASAP Methodology."
- Hart, Sandra G. 2006. "Nasa-Task Load Index (NASA-TLX); 20 Years Later." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (9): 904–8. <https://doi.org/10.1177/154193120605000909>.
- Hart, Sandra G., and Lowell E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." *Advances in Psychology* 52 (C): 139–83. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
- Hause, Matthew. 2011. "'Are We There yet?' Assessing Quality in Model Based Systems Engineering." *21st Annual International Symposium of the International Council on Systems Engineering, INCOSE 2011* 1 (June 2011): 505–17. <https://doi.org/10.1002/j.2334-5837.2011.tb01221.x>.
- Hein, Andreas M., Robert Karban, Tim Weilkens, Michele Zamparelli, and Rudolf Hauber. 2011. *Cookbook for MBSE with SysML*. <https://doi.org/10.13140/2.1.4291.2324>.
- Henderson, Kaitlin, Tom McDermott, Eileen Van Aken, and Alejandro Salado. 2023. "Towards Developing Metrics to Evaluate Digital Engineering." *Systems Engineering* 26 (1): 3–31. <https://doi.org/10.1002/sys.21640>.
- Henderson, Kaitlin, and Alejandro Salado. 2021. "Value and Benefits of Model-Based Systems Engineering (MBSE): Evidence from the Literature." *Systems Engineering* 24 (1): 51–66. <https://doi.org/10.1002/sys.21566>.
- Hernandez, Carlos, and Jose Luis Fernandez-Sanchez. 2017. "Model-Based Systems Engineering to Design Collaborative Robotics Applications." In *2017 IEEE International Symposium on Systems Engineering, ISSE 2017 - Proceedings*. <https://doi.org/10.1109/SysEng.2017.8088258>.
- Herter, Johannes, Ross Brown, and Jivka Ovtcharova. 2013. "A Visual Language for the Collaborative Visualization of Integrated Conceptual Models in Product Development Scenarios." In *Smart Product Engineering*, 805–14. https://doi.org/10.1007/978-3-642-30817-8_79.

- Herzog, Erik, Jessica Hallonquist, and Johan Naeser. 2012. "Systems Modeling with sysML - an Experience Report." In *INCOSE International Symposium, IS*, 1:633–44. <https://doi.org/10.1002/j.2334-5837.2012.tb01359.x>.
- Herzog, Erik, Asmus Pandikow, and Syntell Ab. 2005. "SysML – an Assessment." In *INCOSE International Symposium, IS*.
- Hinton, Perry R., Isabella McMurray, and Charlotte Brownlow. 2014. *SPSS Explained*. Second edition. London ; New York: Routledge, Taylor & Francis Group.
- Hoffmann, Hans-Peter. 2014. "IBM Rational Harmony Deskbook."
- Holladay, Jon B., Jessica Knizhnik, Karen J. Weiland, Amanda Stein, Terry Sanders, and Paul Schwindt. 2019. "MBSE Infusion and Modernization Initiative (MIAMI): 'Hot' Benefits for Real NASA Applications." In *2019 IEEE Aerospace Conference*, 1–14. <https://doi.org/10.1109/AERO.2019.8741795>.
- Holt, Jon. 2021. *Systems Engineering Demystified: A Practitioner's Handbook for Developing Complex Systems Using a Model-Based Approach*. Packt Publishing Ltd.
- Huygelier, Hanne, Brenda Schraepen, Raymond van Ee, Vero Vanden Abeele, and Céline R. Gillebert. 2019. "Acceptance of Immersive Head-Mounted Virtual Reality in Older Adults." *Scientific Reports* 9 (1): 4519. <https://doi.org/10.1038/s41598-019-41200-6>.
- Irani, Pourang, and Colin Ware. 2003. "Diagramming Information Structures Using 3D Perceptual Primitives." *ACM Transactions on Computer-Human Interaction* 10 (1): 1–19. <https://doi.org/10.1145/606658.606659>.
- Isaksson, Ola, Claudia Eckert, Massimo Panarotto, and Johan Malmqvist. 2020. "You Need to Focus to Validate." In *Proceedings of the Design Society: DESIGN Conference*, 1:31–40. Cambridge University Press.
- ISO 9241. 2019. "ISO 9241-210:2019." ISO. 2019. <https://www.iso.org/standard/77520.html>.
- ISO/IEC 19514. 2017. "ISO/IEC 19514:2017." ISO. 2017. <https://www.iso.org/fr/standard/65231.html>.
- "ISO/IEC, ISO/IEC Standard 24744: Software Engineering—Metamodel for Development Methodologies." 2007.
- ISO/IEC/IEEE 15288. 2023. "ISO/IEC/IEEE International Standard - Systems and Software Engineering—System Life Cycle Processes." *ISO/IEC/IEEE 15288:2023(E)*, May, 1–128. <https://doi.org/10.1109/IEEESTD.2023.10123367>.
- ISO/IEC/IEEE 24641. 2023. "ISO/IEC/IEEE International Standard - Systems and Software Engineering—Methods and Tools for Model-Based Systems and Software Engineering." *ISO/IEC/IEEE 24641:2023(E)*, May, 1–98. <https://doi.org/10.1109/IEEESTD.2023.10123376>.
- ISO/IEC/IEEE 29148. 2018. "ISO/IEC/IEEE 29148:2018." ISO. 2018. <https://www.iso.org/standard/72089.html>.
- ISO/IEC/IEEE 42010. 2011. *Systems and Software Engineering—Recommended Practice for Architectural Descriptions of Software-Intensive Systems*.
- . 2022. "IEEE/ISO/IEC International Standard for Software, Systems and Enterprise—Architecture Description." *ISO/IEC/IEEE 42010:2022(E)*, November, 1–74. <https://doi.org/10.1109/IEEESTD.2022.9938446>.
- ISO/IEC/IEEE 42020. 2019. "ISO/IEC/IEEE International Standard - Software, Systems and Enterprise – Architecture Processes." *ISO/IEC/IEEE 42020:2019(E)*, July, 1–126. <https://doi.org/10.1109/IEEESTD.2019.8767004>.

- ISO/IEC/IEEE 42030. 2019. "ISO/IEC/IEEE International Standard - Software, Systems and Enterprise – Architecture Evaluation Framework." *ISO/IEC/IEEE 42030:2019(E)*, July, 1–88. <https://doi.org/10.1109/IEEESTD.2019.8767001>.
- Khatri, Vijay, Iris Vessey, V. Ramesh, Paul Clay, and Sung-Jin Park. 2006. "Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain Knowledge." *Information Systems Research* 17 (1): 81–99.
- Kim, Jinwoo, Jungpil Hahn, and Hyoungmee Hahn. 2000. "How Do We Understand a System with (so) Many Diagrams? Cognitive Integration Processes in Diagrammatic Reasoning." *Information Systems Research*, no. October 2015.
- Kommetter, Christopher, and Martin Ebner. 2019. "A Pedagogical Framework for Mixed Reality in Classrooms Based on a."
- Korfiatis, Peter, Robert Cloutier, and Teresa Zigh. 2012. "Graphical CONOPS Development to Enhance Model Based Systems Engineering." In *Third International Engineering Systems Symposium, CESUN 2012*, 10.
- Kosch, Thomas, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. 2023. "A Survey on Measuring Cognitive Workload in Human-Computer Interaction." *ACM Computing Surveys* 55 (13s): 1–39. <https://doi.org/10.1145/3582272>.
- Krob, Daniel. 2022. *Model-Based Systems Architecting: Using CESAM to Architect Complex Systems*. John Wiley & Sons.
- Krogstie, John, Guttorm Sindre, and Håvard Jørgensen. 2006. "Process Models Representing Knowledge for Action: A Revised Quality Framework." *European Journal of Information Systems* 15 (1): 91–102. <https://doi.org/10.1057/palgrave.ejis.3000598>.
- Kuhar, Saša, and Gregor Polančič. 2021. "Conceptualization, Measurement, and Application of Semantic Transparency in Visual Notations: A Systematic Literature Review." *Software and Systems Modeling* 20 (6): 2155–97. <https://doi.org/10.1007/s10270-021-00888-9>.
- Langeveld, Lau. 2011. "Product Design with Embodiment Design as a New Perspective." In *Industrial Design - New Frontiers*, edited by Denis Coelho. InTech. <https://doi.org/10.5772/20579>.
- Larkin, J.H, and H.A Simon. 1987. "Why a Diagram Is (Sometimes) Worth Ten Thousand Words." *Cognitive Science*.
- Lee, Kwan Min. 2004. "Presence, Explicated." *Communication Theory* 14 (1): 27–50. <https://doi.org/10.1111/j.1468-2885.2004.tb00302.x>.
- Lindland, Odd, Guttorm Sindre, and Arne Sølvberg. 1994. "Understanding Quality in Conceptual Modeling, IEEE Software, 11(2):42-49." *Software, IEEE* 11 (April): 42–49. <https://doi.org/10.1109/52.268955>.
- Lundberg, Mari. 2006. *Kinesiophobia: Various Aspects of Moving with Musculoskeletal Pain*. Göteborg: Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy at Göteborg University.
- Lutfi, Mostafa, and Ricardo Valerdi. 2023. "Integration of SysML and Virtual Reality Environment: A Ground Based Telescope System Example." *Systems* 11 (4): 189. <https://doi.org/10.3390/systems11040189>.
- Mahboob, Atif,) ; Weber, Christian,) ; Husung, Stephan,) ; Liebal, Andreas,) ; Krömker, and Heidi. 2017. "MODEL BASED SYSTEMS ENGINEERING (MBSE) APPROACH FOR CONFIGURABLE PRODUCT USE-CASE SCENARIOS IN VIRTUAL ENVIRONMENTS." In *21ST INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED17*. Vol. 3.

- Maier, Mark W., and Eberhardt Rechtin. 2000. *The Art of Systems Architecting*. 2nd ed. Boca Raton: CRC Press.
- Maletic, J.I., J. Leigh, A. Marcus, and G. Dunlap. 2001. "Visualizing Software in an Immersive Virtual Reality Environment." In *Proceedings 9th International Workshop on Program Comprehension. IWPC 2001*, 26–35. Toronto, Ont., Canada: IEEE Comput. Soc. <https://doi.org/10.1109/WPC.2001.921711>.
- Martinez Escobar, Marisol, Bethany Junke, Joseph Holub, Kenneth Hisley, David Eliot, and Eliot Winer. 2015. "Evaluation of Monoscopic and Stereoscopic Displays for Visual–Spatial Tasks in Medical Contexts." *Computers in Biology and Medicine* 61 (June): 138–43. <https://doi.org/10.1016/j.combiomed.2015.03.026>.
- Masri, Kamal, Drew Parker, and Andrew Gemino. 2008. "Using Iconic Graphics in Entity-Relationship Diagrams: The Impact on Understanding." *Journal of Database Management* 19 (3): 22–41. <https://doi.org/10.4018/jdm.2008070102>.
- Maurandy, Julien, Achim Helm, Eberhard Gill, and Roland Stalford. 2012. "11.5.3 Cost-Benefit Analysis of SysML Modelling for the Atomic Clock Ensemble in Space (ACES) Simulator." *INCOSE International Symposium* 22 (1): 1726–45. <https://doi.org/10.1002/j.2334-5837.2012.tb01433.x>.
- Mayer, Richard E. 2001. *Multimedia Learning*. 2nd edition. Cambridge: Cambridge University Press.
- McDermott, Thomas, Nicole Hutchison, Megan Clifford, Eileen Van Aken, Alejandro Salado, and Kaitlin Henderson. 2020. *Benchmarking the Benefits and Current Maturity of Model-Based Systems Engineering across the Enterprise: Results of the MBSE Maturity Survey*.
- Melo, Miguel, Guilherme Gonçalves, José Vasconcelos-Raposo, and Maximino Bessa. 2023. "How Much Presence Is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score." *IEEE Access* 11: 24675–85. <https://doi.org/10.1109/ACCESS.2023.3254892>.
- Mending, J., H.A. Reijers, and W.M.P. Van Der Aalst. 2010. "Seven Process Modeling Guidelines (7PMG)." *Information and Software Technology* 52 (2): 127–36. <https://doi.org/10.1016/j.infsof.2009.08.004>.
- Mhenni, Faïda, Jean Yves Choley, Olivia Penas, Régis Plateaux, and Moncef Hammadi. 2014. "A SysML-Based Methodology for Mechatronic Systems Architectural Design." *Advanced Engineering Informatics* 28 (3): 218–31. <https://doi.org/10.1016/j.aei.2014.03.006>.
- Milgram, Paul, and Fumio Kishino. 1994. "A Taxonomy of Mixed Reality Visual Displays." *IEICE TRANSACTIONS on Information and Systems* 77 (12): 1321–29.
- Moody, Daniel. 2009. "The Physics of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering." *IEEE Transactions on Software Engineering* 35 (6): 756–79. <https://doi.org/10.1109/TSE.2009.67>.
- Moser, Claus Adolf, and Graham Kalton. 2017. *Survey Methods in Social Investigation*. Routledge.
- Muhanna, Muhanna A. 2015. "Virtual Reality and the CAVE: Taxonomy, Interaction Challenges and Research Directions." *Journal of King Saud University - Computer and Information Sciences* 27 (3): 344–61. <https://doi.org/10.1016/j.jksuci.2014.03.023>.
- Müller, J. 2020. "Does Form Follow Function? Connecting Function Modelling and Geometry Modelling for Design Space Exploration."
- Norman, Donald A. 1988. *The Psychology Of Everyday Things*. New York: Basic Books.

- Oberhauser, Roy. 2021. "VR-UML: The Unified Modeling Language in Virtual Reality – An Immersive Modeling Experience." *Lecture Notes in Business Information Processing* 422 LNBIP (Xmi): 40–58. https://doi.org/10.1007/978-3-030-79976-2_3.
- . 2022. "VR-SysML : SysML Model Visualization and Immersion in Virtual Reality," no. c: 61–66.
- Oberhauser, Roy, and Camil Pogolski. 2019. "VR-EA: Virtual Reality Visualization of Enterprise Architecture Models with ArchiMate and BPMN." In *Business Modeling and Software Design: 9th International Symposium, BMSD 2019, Lisbon, Portugal, July 1–3, 2019, Proceedings 9*, 170–87. Springer.
- Oberhauser, Roy, Camil Pogolski, and Alexandre Matic. 2018. "VR-BPMN: Visualizing BPMN Models in Virtual Reality." In *Business Modeling and Software Design: 8th International Symposium, BMSD 2018, Vienna, Austria, July 2-4, 2018, Proceedings 8*, 83–97. Springer.
- Oberhauser, Roy, Pedro Sousa, and Florian Michel. 2020. "VR-EAT: Visualization of Enterprise Architecture Tool Diagrams in Virtual Reality." In *Business Modeling and Software Design: 10th International Symposium, BMSD 2020, Berlin, Germany, July 6–8, 2020, Proceedings 10*, 221–39. Springer.
- OMG. 2003. "UML for Systems Engineering. Request for Proposal." *Ad/03-03-41*. http://syseng.omg.org/UML_for_SE_RFP.htm.
- . 2019. "OMG Systems Modeling Language (OMG SysML) Specification V1.6."
- Opdahl, Andreas L., and Brian Henderson-Sellers. 2002. "Ontological Evaluation of the UML Using the Bunge–Wand–Weber Model." *Software and Systems Modeling* 1 (1): 43–67. <https://doi.org/10.1007/s10270-002-0003-9>.
- Ozok, A Ant. 2012. "Survey Design and Implementation in HCI."
- Paige, R. F., J. S. Ostroff, and P. J. Brooke. 2000. "Principles for Modeling Language Design." *Information and Software Technology* 42 (10): 665–75. [https://doi.org/10.1016/S0950-5849\(00\)00109-9](https://doi.org/10.1016/S0950-5849(00)00109-9).
- Parrott, Edith, Katie Trase, Randi Green, Denise Varga, and Joe Powell. 2016. "NASA GRC MBSE Implementation Status." February 17. <https://ntrs.nasa.gov/citations/20160012299>.
- Pimentel, Ken. 2000. *Virtual Reality: Through the New Looking Glass*. McGraw-Hill, Inc.
- Piques, J-D, and E Andrianarison. 2012. "SysML for Embedded Automotive Systems: Lessons Learned." In *Embedded Real Time Software & Systems, ERTS*, 2:1–10.
- Piques, Jean-denis. 2014. "SysML for Embedded Automotive Systems: SysCARS Methodology." In *Embedded Real Time Software and Systems (ERTS2014)*. Toulouse, France.
- Popescu, George, and Alain Wegmann. 2014. "Using the Physics of Notations Theory to Evaluate the Visual Notation of SEAM." In *Proceedings - 16th IEEE Conference on Business Informatics, CBI 2014*, 2:166–73. <https://doi.org/10.1109/CBI.2014.21>.
- Poupyrev, Ivan. 1998. "Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques." *Computer Graphics Forum* 17 (3).
- Poupyrev, Ivan, and Tadao Ichikawa. 1999. "Manipulating Objects in Virtual Worlds: Categorization and Empirical Evaluation of Interaction Techniques." *Journal of Visual Languages and Computing*.
- Pugh, S. 1991. "Total Design: Integrated Methods for Successful Product Engineering, S. Pugh, Addison-Wesley, 1990." *Quality and Reliability Engineering International* 7 (2): 119–119. <https://doi.org/10.1002/qre.4680070210>.

- Raimbaud, Pierre. 2020. "Virtual Reality for Building Industry Needs: Guiding the Design of User Interactions through a Task-Centred Methodology."
- Rechtin, Eberhardt. 1991. *Systems Architecting: Creating and Building Complex Systems*. Englewood Cliffs, N.J: Prentice Hall.
- Regenbrecht, Holger, and Thomas Schubert. 2002. "Real and Illusory Interactions Enhance Presence in Virtual Environments." *Presence: Teleoperators & Virtual Environments* 11 (4): 425–34.
- Rick, Adcock, Jackson Scott, Singer Janet, and Hybertson Duane. 2023. "Principles of Systems Thinking - SEBoK." 2023. https://sebokwiki.org/wiki/Principles_of_Systems_Thinking.
- Roozenburg, N.F.M, and J Eekels. 1995. "Product Design: Fundamentals and Methods." 1995. <https://www.abebooks.fr/9780471943518/Product-Design-Fundamentals-Methods-Roozenburg-0471943517/plp>.
- Roques, Pascal. 2016. "MBSE with the ARCADIA Method and the Capella Tool."
- Rumbaugh, James, Ivar Jacobson, and Grady Booch. 1999. *The Unified Modeling Language Reference Manual*. 5. print. The Addison-Wesley Object Technology Series. Reading, Mass.: Addison-Wesley.
- Safran. 2022. "Safran lance la commercialisation d'un logiciel pour optimiser la conception de systèmes complexes." Safran. May 17, 2022. <https://www.safran-group.com/fr/espace-presse/safran-lance-commercialisation-logiciel-optimiser-conception-systemes-complexes-2022-05-17>.
- Sapuan, S. M. 2017. "Chapter 2 - Concurrent Engineering, Product Design, and Development." In *Composite Materials*, edited by S. M. Sapuan, 29–56. Boston: Butterworth-Heinemann. <https://doi.org/10.1016/B978-0-12-802507-9.00002-7>.
- Schreiber, Andreas, and Martin Misiak. 2018. "Visualizing Software Architectures in Virtual Reality with an Island Metaphor." In *Virtual, Augmented and Mixed Reality, VAMR*, 1:168–82.
- Schreiber, Andreas, Lisa Nafeie, Artur Baranowski, Peter Seipel, and Martin Misiak. 2019. "Visualization of Software Architectures in Virtual Reality and Augmented Reality." In *IEEE Aerospace Conference*.
- Schubert, Thomas, Frank Friedmann, and Holger Regenbrecht. 2001. "The Experience of Presence: Factor Analytic Insights." *Presence: Teleoperators and Virtual Environments* 10 (3): 266–81. <https://doi.org/10.1162/105474601300343603>.
- Schuette, Reinhard, and Thomas Rotthowe. 1998. "The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models." In *Conceptual Modeling – ER '98*, edited by Tok-Wang Ling, Sudha Ram, and Mong Li Lee, 1507:240–54. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-49524-6_20.
- Shaft, Teresa M., and Iris Vessey. 1998. "The Relevance of Application Domain Knowledge: Characterizing the Computer Program Comprehension Process." *Journal of Management Information Systems* 15 (1): 51–78. <https://doi.org/10.1080/07421222.1998.11518196>.
- Siau, Keng, and Qing Cao. 2001. "Unified Modeling Language: A Complexity Analysis." *Journal of Database Management (JDM)* 12 (1): 26–34. <https://doi.org/10.4018/jdm.2001010103>.
- Simon, Herbert A. 1978. "On the Forms of Mental Representation."

- Simon, Herbert A., and Glenn Lea. 1974. "Problem Solving and Rule Induction: A Unified View." In *Knowledge and Cognition*, ix, 321–ix, 321. Oxford, England: Lawrence Erlbaum.
- Slater, Mel, Vasilis Linakis, Martin Usoh, and Rob Kooper. 1999. "Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess."
- Slater, Mel, and Anthony Steed. 2000. "A Virtual Presence Counter." *Presence* 9 (5): 413–34.
- Slater, Mel, and Sylvia Wilbur. 1997. "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments." *Presence: Teleoperators & Virtual Environments* 6 (6): 603–16.
- Sollenberger, Randy L., and Paul Milgram. 1993. "Effects of Stereoscopic and Rotational Displays in a Three-Dimensional Path- Tracing Task." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 35 (3): 483–99. <https://doi.org/10.1177/001872089303500306>.
- Stanney, K. 1995. "Realizing the Full Potential of Virtual Reality: Human Factors Issues That Could Stand in the Way." In *Proceedings Virtual Reality Annual International Symposium '95*, 28–34. Research Triangle Park, NC, USA: IEEE Comput. Soc. Press. <https://doi.org/10.1109/VRAIS.1995.512476>.
- Storrie, Harald. 2013. "Towards an Operationalization of the Physics of Notation for the Analysis of Visual Languages" 8107 (September). <https://doi.org/10.1007/978-3-642-41533-3>.
- Straub, Detmar, Marie-claude Boudreau, and David Gefen. 2004. "Validation Guidelines for IS Positivist Research." *Communications of the Association for Information Systems* 3 (January). <https://doi.org/10.17705/1CAIS.01324>.
- Sutherland, Ivan E. 1965. "The Ultimate Display." In *Proceedings of the IFIP Congress*, 2:506–8. New York.
- Sweller, John. 1988. "Cognitive Load during Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–85. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7).
- Sweller, John, and Paul Chandler. 1994. "Why Some Material Is Difficult to Learn." *Paul Chandler* 12 (September). https://doi.org/10.1207/s1532690xci1203_1.
- Sweller, John, Jeroen J. G. Van Merriënboer, and Fred G. W. C. Paas. 1998. "Cognitive Architecture and Instructional Design." *Educational Psychology Review* 10 (3): 251–96. <https://doi.org/10.1023/A:1022193728205>.
- Taherdoost, Hamed. 2016. "Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3205040>.
- Thalheim, Bernhard. 2009. "Towards a Theory of Conceptual Modelling." In *Advances in Conceptual Modeling - Challenging Perspectives*, edited by Carlos Alberto Heuser and Günther Pernul, 5833:45–54. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04947-7_7.
- . 2018. "Conceptual Model Notions – A Matter of Controversy: Conceptual Modelling and Its Lacunas." *Enterprise Modelling and Information Systems Architectures*, February, 9-27 Pages. <https://doi.org/10.18417/EMISA.SI.HCM.1>.
- Tran, Thi, Gilles Foucault, and Romain Pinquie. 2021. "Benchmarking of 3D Modelling in Virtual Reality." In *CAD'21 Proceedings*, 324–28. CAD Solutions LLC. <https://doi.org/10.14733/cadconfP.2021.324-328>.

- Van Der Linden, Dirk, and Irit Hadar. 2019. "A Systematic Literature Review of Applications of the Physics of Notations." *IEEE Transactions on Software Engineering* 45 (8): 736–59. <https://doi.org/10.1109/TSE.2018.2802910>.
- Van Der Linden, Dirk, Anna Zamansky, and Irit Hadar. 2016. "How Cognitively Effective Is a Visual Notation? On the Inherent Difficulty of Operationalizing the Physics of Notations." In *Enterprise, Business-Process and Information Systems Modeling*, edited by Rainer Schmidt, Wided Guédria, Iliá Bider, and Sérgio Guerreiro, 248:448–62. Lecture Notes in Business Information Processing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39429-9_28.
- Vermaas, Pieter E. 2014. "Design Theories, Models and Their Testing: On the Scientific Status of Design Research." In *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations*, edited by Amaresh Chakrabarti and Lucienne T. M. Blessing, 47–66. London: Springer. https://doi.org/10.1007/978-1-4471-6338-1_2.
- Vincur, Juraj, Pavol Návrát, and Ivan Polasek. 2017. "Vr City: Software Analysis in Virtual Reality Environment." In *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 509–16. IEEE.
- Visser, Willemien. 2010. "Visser: Design as Construction of Representations." *Collection*, no. 2: 29–43.
- Voirin, Jean-Luc, ed. 2018. *Model-Based System and Architecture Engineering with the Arcadia Method Bibliography*. Elsevier. <https://doi.org/10.1016/B978-1-78548-169-7.50030-5>.
- Wai-keung Fung, Wang-tai Lo, Yun-hui Liu, and Ning Xi. 2005. "A Case Study of 3D Stereoscopic vs. 2D Monoscopic Tele-Reality in Real-Time Dexterous Teleoperation." In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 181–86. Edmonton, Alta., Canada: IEEE. <https://doi.org/10.1109/IROS.2005.1545299>.
- Walden, David D., Garry J. Roedler, Kevin Forsberg, R. Douglas Hamelin, Thomas M. Shortell, and International Council on Systems Engineering. 2015. *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*.
- Wassermann, Benjamin, Tino Bog, Stefan Kollmannsberger, and Ernst Rank. 2016. "A DESIGN-THROUGH-ANALYSIS APPROACH USING THE FINITE CELL METHOD." In *Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2016)*, 2601–13. Crete Island, Greece: Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece. <https://doi.org/10.7712/100016.1984.8920>.
- Weilkiens, Tim. 2020. *SYSMOD - The Systems Modeling Toolbox*. Edited by MBSE4U.
- Wickens, Christopher D., Sallie E. Gordon, Yili Liu, and J. Lee. 2004. *An Introduction to Human Factors Engineering*. Vol. 2. Pearson Prentice Hall Upper Saddle River, NJ.
- Witmer, Bob G., and Michael J. Singer. 1998. "Measuring Presence in Virtual Environments: A Presence Questionnaire." *Presence* 7 (3): 225–40.
- Witrock, Merlin C. 1989. "Generative Processes of Comprehension." *Educational Psychologist* 24 (4): 345–76. https://doi.org/10.1207/s15326985ep2404_2.
- Woods, David D., and Jennifer C. Watts. 1997. "Chapter 26 - How Not to Have to Navigate Through Too Many Displays." In *Handbook of Human-Computer Interaction (Second Edition)*, edited by Marting G. Helander, Thomas K. Landauer, and Prasad V. Prabh, 617–50. Amsterdam: North-Holland. <https://doi.org/10.1016/B978-044481862-1.50092-3>.

- Yigitbas, Enes, Simon Gorissen, Nils Weidmann, and Gregor Engels. 2022. "Design and Evaluation of a Collaborative UML Modeling Environment in Virtual Reality." *Software and Systems Modeling*. <https://doi.org/10.1007/s10270-022-01065-2>.
- Younse, Paulo J., Jessica E. Cameron, and Thomas H. Bradley. 2021. "Comparative Analysis of a Model-Based Systems Engineering Approach to a Traditional Systems Engineering Approach for Architecting a Robotic Space System through Knowledge Categorization." *Systems Engineering* 24 (3): 177–99. <https://doi.org/10.1002/sys.21573>.
- Ziehmann, J., and B. Lantow. 2021. "Moody's Physics of Notations: High Impact, Little Support." *CEUR Workshop Proceedings* 3045: 31–38.

APPENDICES

Appendix A Systems thinking principles (Rick et al. 2023)

Table 76 Systems Principles (Rick et al. 2023)

Abstraction	Focusing on essential characteristics is important in problem solving because it allows problem solvers to ignore the nonessential, thus simplifying the problem.
Boundary	A boundary or membrane separates the system from the external world. It serves to concentrate interactions inside the system while allowing exchange with external systems.
Change	Change is necessary for growth and adaptation and should be accepted and planned for as part of the natural order of things rather than something to be ignored, avoided, or prohibited.
Dualism	Recognize dualities and consider how they are, or can be, harmonized in the context of a larger whole.
Encapsulation	Hide internal parts and their interactions from the external environment.
Equifinality	In open systems, the same final state may be reached from different initial conditions and in different ways. This principle can be exploited, especially in systems of purposeful agents.
Holism	A system should be considered as a single entity, a whole, not just as a set of parts.
Interaction	The properties, capabilities, and behavior of a system are derived from its parts, from interactions between those parts, and from interactions with other systems.
Layer Hierarchy	The evolution of complex systems is facilitated by their hierarchical structure (including stable intermediate forms) and the understanding of complex systems is facilitated by their hierarchical description.
Leverage	Achieve maximum leverage. Because of the power versus generality tradeoff, leverage can be achieved by a complete solution (power) for a narrow class of problems, or by a partial solution for a broad class of problems (generality).
Modularity	Unrelated parts of the system should be separated, and related parts of the system should be grouped together.

Network	The network is a fundamental topology for systems that forms the basis of togetherness, connection, and dynamic interaction of parts that yield the behavior of complex systems.
Parsimony	One should choose the simplest explanation of a phenomenon, the one that requires the fewest assumptions. This applies not only to choosing a design, but also to operations and requirements.
Regularity	Systems science should find and capture regularities in systems, because those regularities promote systems understanding and facilitate systems practice.
Relations	A system is characterized by its relations: the interconnections between the elements. Feedback is a type of relation. The set of relations defines the network of the system.
Separation of Concerns	A larger problem is more effectively solved when decomposed into a set of smaller problems or concerns.
Similarity/Difference	Both the similarities and differences in systems should be recognized and accepted for what they are. Avoid forcing one size fits all, and avoid treating everything as entirely unique.
Stability/Change	Things change at different rates, and entities or concepts at the stable end of the spectrum can and should be used to provide a guiding context for rapidly changing entities at the volatile end of the spectrum. The study of complex adaptive systems can give guidance to system behavior and design in changing environments.
Synthesis	Systems can be created by “choosing (conceiving, designing, selecting) the right parts, bringing them together to interact in the right way, and in orchestrating those interactions to create requisite properties of the whole, such that it performs with optimum effectiveness in its operational environment, so solving the problem that prompted its creation”.
View	Multiple views, each based on a system aspect or concern, are essential to understand a complex system or problem situation. One critical view is how concern relates to properties of the whole.

Appendix B Tasks from ISO 15288

B.1 Business or mission analysis process

Table 77 Activities and tasks of business or mission analysis process (ISO/IEC/IEEE 15288 2023)

A1 Prepare for business or mission analysis.	
T11	Review changes to the organization strategy and concept of operations to identify potential problems and opportunities with respect to desired organization mission(s), vision, goals, and objectives.
T12	Define the business or mission analysis strategy.
T13	Identify and plan for the necessary enabling systems or services needed to support business or mission analysis.
T14	Obtain or acquire access to the enabling systems or services to be used.
A2 Define the problem or opportunity space.	
T21	Analyse the problems and opportunities in the context of relevant trade-space factors.
T22	Define the mission, business, or operational problem or opportunity to be addressed by a solution.
T23	Prioritise the potential problem or opportunity against other business needs.
A3 Characterize the solution space. This activity consists of the following tasks.	
T31	Define preliminary operational concepts and other life cycle concepts.
T32	Identify alternative solution classes that span the potential solution space.
A4 Evaluate alternative solution classes.	
T41	Assess each alternative solution class.
T42	Select the preferred alternative solution class(es).
T43	Provide feedback to strategic level life cycle concepts to reflect the selected solution class(es).
A5 Manage the business or mission analysis.	
T51	Record key business or mission analysis decisions and the rationale.
T52	Maintain traceability of business or mission analysis and the alternative solution class(es).
T53	Provide key artefacts that have been selected for baselines.

B.2 Stakeholder needs and requirements definition process

Table 78 Activities and tasks of stakeholder needs and requirements definition process (ISO/IEC/IEEE 15288 2023)

A1 Prepare for stakeholder needs and requirements definition.	
T11	Identify the stakeholders who have an interest in the solution throughout its life cycle.
T12	Define the stakeholder needs and requirements definition strategy.
T13	Identify and plan for the necessary enabling systems or services needed to support stakeholder needs and requirements definition.

T14	Obtain or acquire access to the enabling systems or services to be used.
A2 Develop the operational concept and other life cycle concepts.	
T21	Define context of use within the concept of operations, the preliminary life cycle concepts, and the preferred solution class(es).
T22	Define the context of use and a set of scenarios (or use cases) to identify all required capabilities that correspond to anticipated operational concepts and other life cycle concepts.
T23	Characterize the operational environment and the intended users.
T24	Identify interactions between users and the system and the factors affecting the interactions.
T25	Identify all interface boundaries across which the Sol interacts with external systems.
T26	Identify the constraints on a system solution.
A3 Define stakeholder needs.	
T31	Identify stakeholder needs within the constraints imposed by the life cycle concepts.
T32	Prioritise and down-select needs.
T33	Record the stakeholder needs and rationale.
A4 Transform stakeholder needs into stakeholder requirements.	
T41	Identify the stakeholder requirements and functions that relate to critical quality characteristics, such as assurance, safety, security, environment, or health.
T42	Define stakeholder requirements, consistent with life cycle concepts, scenarios, interactions, constraints, critical quality characteristics, and SoS considerations.
A5 Analyse stakeholder needs and requirements.	
T51	Analyse the complete set of stakeholder requirements.
T52	Define critical performance measures and quality characteristics that enable the assessment of technical achievement.
T53	Feed back the analysed requirements to applicable stakeholders to validate that their needs and expectations have been adequately captured and expressed.
T54	Resolve stakeholder requirements issues.
A6 Manage the stakeholder needs and requirements definition.	
T61	Obtain explicit agreement on the stakeholder requirements.
T62	Record key stakeholder requirements decisions and the rationale.
T63	Maintain traceability of stakeholder needs and requirements.
T64	Provide key artefacts that have been selected for baselines

B.3 System requirements definition process

Table 79 Activities and tasks of system requirements definition process (ISO/IEC/IEEE 15288 2023)

A1 Prepare for system requirements definition.	
T11	Define the functional boundary of the system in terms of the behaviour and properties to be provided.
T12	Define the system requirements definition strategy.
T13	Identify and plan for the necessary enabling systems or services needed to support system requirements definition.

T14	Obtain or acquire access to the enabling systems or services to be used.
A2 Define system requirements.	
T21	Define each function that the system is required to perform.
T22	Define necessary implementation constraints.
T23	Identify system requirements that relate to risks, criticality of the system, or critical quality characteristics.
T24	Define system requirements and rationale.
A3 Analyse system requirements.	
T31	Analyse the complete set of system requirements.
T32	Define critical performance measures that enable the assessment of technical achievement.
T33	Feed back the analysed requirements to applicable stakeholders for review.
T34	Resolve system requirements issues.
A4 Manage system requirements.	
T41	Obtain explicit agreement on the system requirements.
T42	Record key system requirements decisions and the rationale.
T43	Maintain traceability of the system requirements.
T44	Provide key artefacts that have been selected for baselines.

B.4 System architecture definition process

Table 80 Activities and tasks of system architecture definition process (ISO/IEC/IEEE 15288 2023)

A1 Prepare for system architecture definition.	
T11	Identify key milestones and decisions to be informed by the system architecture effort.
T12	Define the strategy for system architecture definition.
T13	Prepare for and plan the support to architecture governance and architecture management efforts of the organisation.
T14	Identify and plan for the necessary enabling systems or services needed to support system architecture definition efforts.
T15	Obtain or acquire access to the enabling systems or services to be used in the system architecture definition efforts
A2 Conceptualise the system architecture.	
T21	Characterize the problem space.
T22	Establish architecture objectives and critical success criteria.
T23	Synthesize potential solution(s) in the solution space.
T24	Characterize solutions and the trade space.
T25	Formulate candidate architecture(s).
T26	Capture architecture concepts and properties.
T27	Relate the architecture to other architectures and to relevant affected entities to help ensure consistency.
T28	Coordinate use of architecture by intended users.
A3 Evaluate the system architecture.	

T31	Determine evaluation objectives and criteria.
T32	Determine evaluation methods and integrate with evaluation objectives and criteria.
T33	Collect and review evaluation-related information.
T34	Analyse architecture concepts and properties and assess the value of the architecture.
T35	Combine the analyses and assessments into an overall evaluation to select a preferred system architecture solution.
T36	Characterize architecture(s) based on assessment results.
T37	Formulate findings and recommendations.
T38	Capture and communicate evaluation results.
A4 Elaborate the system architecture.	
T41	Identify or develop architecture viewpoints and model kinds and legends that are governed by these architecture viewpoints.
T42	Develop models and views of the architecture(s).
T43	Relate the architecture to other architectures and to relevant affected entities to help ensure consistency of the elaborated system architecture.
T44	Assess the architecture elaboration.
T45	Coordinate use of elaborated architecture by intended users.
A5 Manage results of system architecture definition.	
T51	Monitor, assess, and control the system architecture definition activities and tasks.
T52	Obtain agreement on the architecture definition
T53	Provide support to organizational architecture governance and architecture management efforts
T54	Record key system architecture decisions and the rationale.
T55	Maintain traceability of the system architecture.
T56	Provide key artefacts that have been selected for baselines.

Appendix C Tasks from ISO 24641

C.1 Produce system models process

Table 81 Activities and tasks of produce system models process (ISO/IEC/IEEE 24641 2023)

A1 Collect engineering data.	
T11	Collect required data from the stakeholders.
A2 Build descriptive models.	
T21	Identify and analyse the concepts, their relationships, and the structure of the system-of-interest.
T22	Identify and formalize the operational needs and missions.
T23	Identify and formalize the system environment, capabilities, boundary or interfaces.
T24	Identify and formalize the system logical structure or architecture.
T25	Refine the system capabilities according to its logical structure or architecture.
T26	Identify and formalize the candidate system organic or physical structures or architectures.
T27	Identify and formalize the breakdown structure (i.e. identify configuration items (CIs)) with reference to the integration strategy (i.e. integration level of responsibility, CIs time of deliveries, etc.).
A3 Build analytical models.	
T31	Produce the analytical models.
T32	Analyse the system structure or architecture static and dynamic or behavioural properties (formalized in the descriptive models) regarding operational needs and missions, specified performances and MBSSE modelling goals.
T33	Identify the best architectural compromise or trade-off taking into account all system aspects, including safety, mass, cost, etc.

C.2 Perform business and mission analysis process

Table 82 Activities and tasks of perform business and mission analysis process (ISO/IEC/IEEE 24641 2023)

A1 Describe high-level target enterprise architectures using models.	
T11	Create goals, associated benefits, and required capabilities.
T12	Link goals and capabilities.
T13	Define the capability phasing.
T14	Establish candidate architectures consisting of different technologies and services (i.e. variant modelling)
T15	Model mission contexts, phases, scenarios and associated mission success criteria.
T16	Model mission environment
A2 Evaluate candidate architectures and analyse gaps using models.	
T21	Perform trade-off between candidate architectures.
T22	Analyse gaps between existing and target architectures.
A3 Establish capability roadmaps.	

T31	Organize captured capabilities into roadmaps with provisioning technologies and services, by taking into account schedule, budget, resource constraints.
T32	Share and updating roadmaps.
A4 Define business and mission requirements.	
T41	Create the requirements and their traceability to the corresponding model elements.
A5 Generate ConOps.	
T51	Define graphical and textual model contents using a meta-model.
T52	Manage the document structure and linked model elements within a repository.

C.3 Perform operational analysis process

Table 83 Activities and tasks of perform operational analysis process (ISO/IEC/IEEE 24641 2023)

A1 Identify system life cycle, boundary and context.	
T11	Model system life cycle.
T12	Model system context by defining the boundary, elements within environment external to the system-of-interest and the interaction with those elements.
T13	Realize preliminary life cycle concept (OpsCon, utilization concept, disposal concept, etc.) models.
A2 Identify stakeholders.	
T21	Discovery and involve stakeholders.
T22	Interview stakeholders.
T23	Capture stakeholder interests.
T24	Classify stakeholders.
A3 Identify use cases and develop use case scenarios, validation scenarios.	
T31	Produce use case diagrams for each life cycle stage according to the modelling objectives (utilization, maintenance, disposal, etc.).
T32	Analyse, refine and complete operations identified and modelled within mission scenarios.
T33	Model use case scenarios, in the way that they can be used for early validation (execution or simulation) and to support validation plans.
A4 Identify operational modes.	
T41	Model operational modes and transitions (i.e. state machines).
T42	Link the scenarios to operational modes.
A5 Capture stakeholder requirement and measures of effectiveness (MOEs).	
T51	Analyse stakeholder needs and constraints using operational models.
T52	Capture formalized stakeholder requirements.
T53	Define and capturing MOEs.

C.4 Perform functional analysis process

Table 84 Activities and tasks of perform functional analysis process (ISO/IEC/IEEE 24641 2023)

A1 Realize functional analysis and decomposition.	
T11	Extract system services from use case (UC) scenarios and identify required high-level system functions.
T12	Decompose them using appropriate functional analysis techniques such as function analysis system technique (FAST) or functional flow block diagram (FFBD).
A2 Detect or identify possible dysfunctions.	
T21	Define quality objectives as safety and reliability.
T22	Identify system dysfunctions from use case alternate flows and what-if scenarios.
T23	Produce dysfunctional and control models.
A3 Identify system elements and allocate functions and develop functional flows.	
T31	Develop functional flows
T32	Develop system state transitions and active functional flows within different states.
A4 Capture system requirements, constraints and measure of performance (MOPs).	
T41	Formalize system functional requirements.
T42	Formalize system interface requirements.
T43	Formalize system performance and quality requirements.
T44	Capture measures of performances (MOPs) and other technical measures.
T45	Define verification means.
A5 Realize and manage traceability.	
T51	Trace system operational activity to functions.
T52	Trace stakeholder requirements to system requirements.

C.5 Perform system structure design process

Table 85 Activities and tasks of perform system structure design process (ISO/IEC/IEEE 24641 2023)

A1 Realize system logical structure.	
T11	Realize system decomposition (system breakdown structure -SBS).
T12	Define subsystems communication or functional interfaces.
T13	Identify alternatives and realize trade-offs.
T14	Refine models.
A2 Realize system physical structure.	
T21	Define physical structures implementing logical systems and system elements.
T22	Identify alternatives and realize trade-offs.
T23	Define product breakdown structure (PBS) and produce CIs, bill of materials, etc.
T24	Provide the models to HFE for HFE analysis, if required.
T25	Refine models.
A3 Realize and manage traceability.	
T31	Trace logical system elements to physical elements.
T32	Trace, make or buy components with the corresponding requirements.

Appendix D Demographic Questions

- What is your level of experience in using conceptual modelling notations for Model-Based Systems Engineering (e.g., SysML, Capella DSL, OPM DSL, System Composer DSL, etc.):
 - **Novice:** I never heard about MBSE languages.
 - **MBSE Model User:** I am able to interpret and understand MBSE models of a system.
 - **SysML Model Builder Fundamental:** I am able to construct a basic MBSE model.
 - **SysML Model Builder Intermediate:** I am able to develop system models using the full MBSE feature set enabling more detailed and precise modelling. I am able to assess aspects of model quality including conformance to language and method constraints.
 - **SysML Model Builder Advanced:** I am able to establish and support a system modelling culture and environment, leveraging the capabilities of MBSE across the organization. I am able to tailor a MBSE modelling language and software to an existing new product development process.
- What is your level of experience in systems architecting (in this survey, systems architecting concentrates on the definition of the system lifecycle/context/stakeholders/interfaces/functions/requirements and the recursive process to decompose a system into sub-systems)?
 - **Novice:** I have never been involved in the definition of a system architecture.
 - **Awareness:** I have been involved in the definition of a system architecture on a toy problem.
 - **Supervised practitioner:** I have been involved in the definition of system architectures on industrial problems, but I require guidance and supervision.
 - **Practitioner:** I have participated in the architecture definition of several industrial systems. I am capable of guiding and advising others.
 - **Expert:** I have an extensive and substantial practical experience in systems architecting. I have defined or tailored processes, methods and tools for architecture definition within organizations. I can teach system architecting.
- What is your level of experience in Systems Engineering (in this survey, Systems Engineering relates to the technical processes defined in the ISO 15288, ISO 42010, ISO 42020, ARP 4754A, INCOSE SE Handbook, NASA SE handbook...)?
 - **Novice:** I have never heard about the technical processes of the systems engineering approach.
 - **Awareness:** I am aware of the systems engineering approach and have applied systems engineering fundamentals to toy problems for educational purposes.
 - **Supervised practitioner:** I understand the fundamentals of the systems engineering approach and have applied them in a professional environment, but I require guidance and supervision.
 - **Practitioner:** I have detailed knowledge of the systems engineering approach. I have contributed to several projects adopting the systems engineering approach. I am capable of guiding and advising others desiring to apply systems engineering fundamentals.
 - **Expert:** I have extensive and substantial practical experience in systems engineering. I have created a systems engineering training course and trained

people on the systems engineering approach. I have defined or tailored my organization's meta-model, processes, systems engineering method, modelling methods, and software.

- What is your level of experience in Virtual Reality:
 - **Novice:** I have never experienced any immersive virtual environment.
 - **VR Awareness:** I have briefly experienced an immersive virtual environment.
 - **VR User:** I often use virtual reality but never developed any applications.
 - **VR Builder Fundamental:** I have developed basic virtual reality applications for educational purposes.
 - **VR Builder Advanced:** I have developed advanced virtual reality applications to solve industrial problems.
 - **VR Experts:** I can teach advanced topics related to virtual reality and how to develop virtual reality applications.

Appendix E User Satisfaction Questions

E.1 SUS Questionnaire

1: I think that I would like to use this modelling interface frequently.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

2: I found the modelling interface unnecessarily complex.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

3: I thought the modelling interface was easy to use.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

4: I think that I would need the support of a technical person to be able to use this modelling interface.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree

- Strongly Agree

5: I found the various functions in this modelling interface were well integrated.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

6: I thought there was too much inconsistency in this modelling interface.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

7: I would imagine that most people would learn to use this modelling interface very quickly.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

8: I found the modelling interface very cumbersome to use.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree

- Strongly Agree

9: I felt very confident using the modelling interface.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

10: I needed to learn a lot of things before I could get going with this modelling interface.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

E.2 Complementary Questions

1: The modelling interface is very efficient to avoid wasting efforts and time while performing a task

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

2: The modelling interface is very effective in producing the intended results.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

3: The modelling interface was helpful to remember the elements (external entities, interfaces, functions, functional scenarios...) of your system architecture.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

4: I think I will use a similar modelling interface for performing model-based systems architecture activities in the future.

- Strongly Disagree
- Disagree

- Neither Agree nor Disagree
- Agree
- Strongly Agree

5: I felt having fun when using the modelling interface.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

6: I think the modelling interface provides a high level of motivation.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

Appendix F Visual Notation Questionnaire

1: The graphical objects are:

- Extremely easy to differentiate
- Easy to differentiate
- Neither easy nor difficult to differentiate
- Difficult to differentiate
- Extremely difficult to differentiate

2: The graphical objects can be confused with each other:

- Very easily
- Fairly easily
- Occasionally
- Rarely
- Never

3: The extraction of meaning from the graphical objects is:

- Extremely intuitive
- Very intuitive
- Moderately intuitive
- Slightly intuitive
- Not intuitive

4: Even without explanation, what the graphical objects represent is:

- Extremely clear
- Very clear
- Moderately clear
- Slightly clear
- Not clear

5: The number of graphical objects is:

- Extremely high
- Quite high
- Neither too high nor too low
- Somewhat low
- Very low

6: The complete representation formed by graphical objects is:

- Extremely difficult to understand
- Very difficult to understand
- Moderately difficult to understand
- Slightly difficult to understand
- Easy to understand

7: The visual expressiveness of the graphical objects can be improved by:

- Significantly reducing variation of color, size, form or brightness
- Reducing variation of color, size, form or brightness
- No need for improvement
- Increasing variation of color, size, form or brightness
- Significantly increasing variation of color, size, form or brightness

8: I perceive the graphical objects as:

- Extremely visually expressive
- Very visually expressive
- Moderately visually expressive
- Slightly visually expressive
- Not visually expressive

9: The complete representation can be improved by:

- Significantly reducing the diversity of graphical symbols
- Reducing the diversity of graphical symbols
- No need for improvement

- Increasing the diversity of graphical symbols
- Significantly increasing the diversity of graphical symbols

10: The number of different graphical objects is:

- Excessively large
- Quite large
- Neither too large nor too small
- Somewhat small
- Very small

Appendix G Igroup Presence Questionnaire

1: How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

extremely aware not aware at all
 -3 -2 -1 0 +1 +2 +3

2: How real did the virtual world seem to you?

completely real not real at all
 -3 -2 -1 0 +1 +2 +3

3: I had a sense of acting in the virtual space, rather than operating something from outside.

fully disagree fully agree
 -3 -2 -1 0 +1 +2 +3

4: How much did your experience in the virtual environment seem consistent with your real world experience?

not consistent very consistent
 -3 -2 -1 0 +1 +2 +3

5: How real did the virtual world seem to you?

about as real as indistinguishable
 an imagined -3 -2 -1 0 +1 +2 +3 from the real
 world world

6: I did not feel present in the virtual space.

did not feel felt present
 -3 -2 -1 0 +1 +2 +3

7: I was not aware of my real environment.

fully disagree fully agree
 -3 -2 -1 0 +1 +2 +3

8: In the computer generated world, I had a sense of "being there"

not at all very much
 -3 -2 -1 0 +1 +2 +3

9: Somehow I felt that the virtual world surrounded me.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

10: I felt present in the virtual space.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

11: I still paid attention to the real environment.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

12: The virtual world seemed more realistic than the real world.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

13: I felt like I was just perceiving pictures.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

14: I was completely captivated by the virtual world.

fully disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fully agree
	-3	-2	-1	0	+1	+2	+3	

Appendix I Group Dynamic Questions

1: The communication between group members was efficient to reach a consensus.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

2: The communication between group members was effective to produce the intended results.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

3: I felt motivated to participant in the conversation.

- Strongly Disagree
- Disagree
- Neither Agree nor Disagree
- Agree
- Strongly Agree

Appendix J Single-Person Controlled Experiment Slides



Immersive Human-Computer Interface for Modelling Systems Architectures: Single-Person Controlled Experimentation

Haobo Wang & Romain Pinqu  – Haobo.wang@grenoble-inp.fr



Motivations & Objectives

Motivations

- Powerfull MBSE modelling interfaces, but with a steep learning curve (abstract representations, limited modelling space, etc.)
- What could be the future MBSE modelling interfaces in 5 to 10 years?

Researchobjective

- Our research studies concentrate on new interactive interfaces for modelling the architecture of engineered systems.
- To conduct exploratory and formative research to improve human-model interactions rather than discover "the scientific truth".

Singlepersoncontrolledexperimentationobjective

- To evaluate the immersive modelling interface from the following perspectives:
 - User satisfaction
 - Visual notation
 - Presence
 - Cognitive load

Single-person controlled experimentation Protocol

1. Consent form
2. Introduction to the mission of an Electric Toothbrush System
3. List of system architecture tasks to achieve
4. Training with Telescope system (15')
5. Exercise with Electric Toothbrush System (25')
6. Post-experimentation questionnaire (15')

3



Introduction



System Mission

What? The electric toothbrush system has the mission is to clean users teeth. The electric toothbrush will also monitor brushing data so as to increase the brushing efficiency by sending recommendations to the user through a mobile application. The electric toothbrush system will be connected to collect the following data: Start time; End time; % of time used for each mode; and Time spent with a brushing pressure exceeding the threshold specified by healthcare recommendations.

Why? To rapidly (approximately between 2 and 3 minutes) clean users teeth.

Who? Teenagers and adults from 14 to 99 years old.

Where? At any place (home, airport, train station, hotel, etc.) but mainly in the bathroom.

When? Three times per day (after breakfast, lunch and dinner).



5

Tasks

As a system architect, you will partially develop the architecture of a Connected Electric Toothbrush System. The scope of the system analysis will be limited to the system level and pave the way to the definition of system requirements.

During the individual session, you must achieve 4 tasks:

1. Define the life cycle of the Connected Electric Toothbrush System
2. Define the external entities for a context of use (e.g., at home) of the Connected Electric Toothbrush System
 - Stakeholders
 - External systems
3. For a context of use (e.g., at home), define the external interfaces (ports and flows) between the Connected Electric Toothbrush System and the external entities.
4. **(extra task)** For the context of use (e.g., at home), define functional scenario (functional chains) of the Connected Electric Toothbrush System.
 - Scenario : To brush teeth

6



Training



Training

There will be several tutorial videos to demonstrate the different interactions in the virtual environment. Most of the interactions are done with the **triggerbutton** on the right controller if there is no specification for the interaction button. Once you finish each video, you can try to repeat what is shown in the video yourself in the virtual environment.

(Please don't use the function **Delete** in the environment, when it's redundant, you can always drag and drop it somewhere that doesn't bother you)

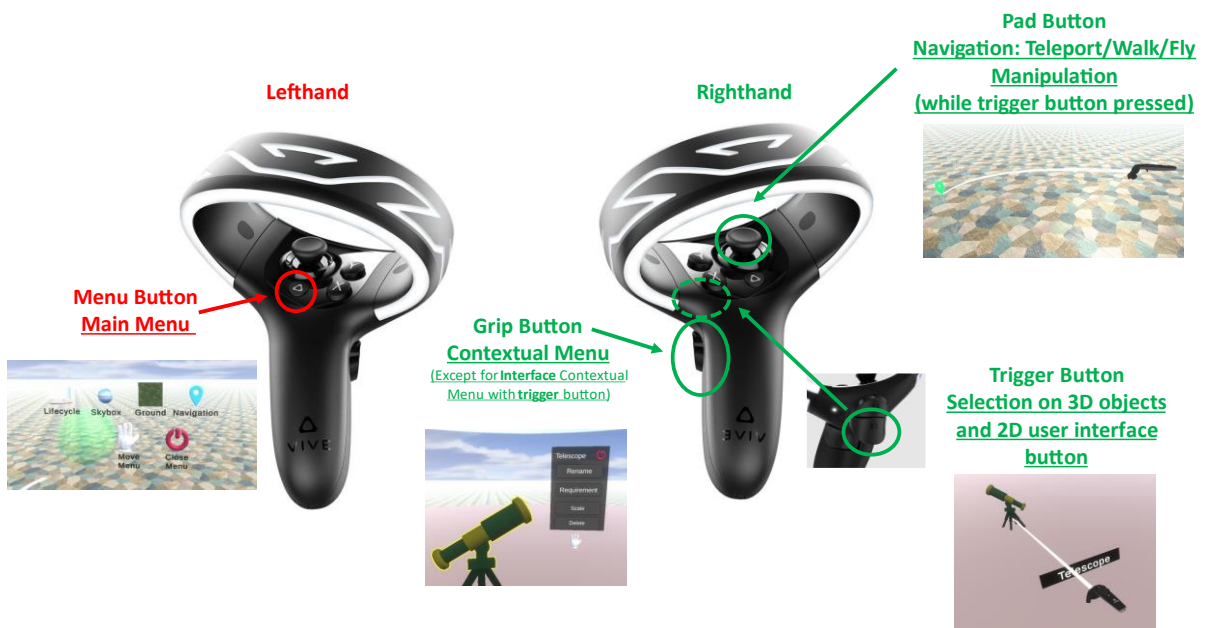
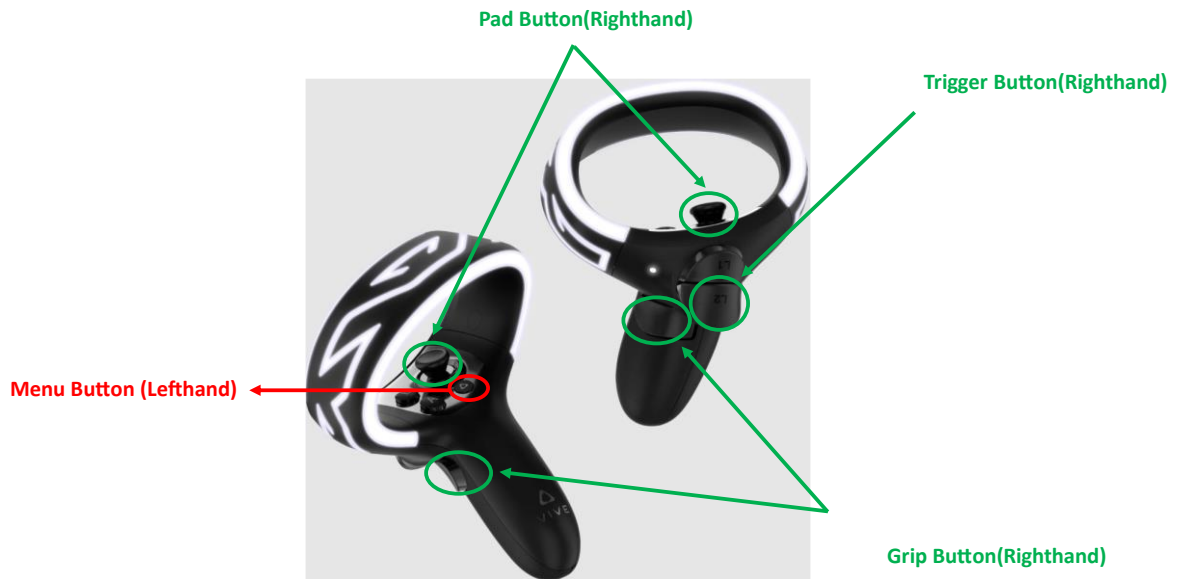
Redo the same process until you finish all the tutorial videos, then you can start the toothbrush exercise.

(Before you start the toothbrush exercise please call Camilo to start recording your screen)

4 types of button to be used:



4 types of button to be used:



Teleport

1. Please watch the tutorial 1-Teleport(D:\Data\Haob\ Sujet)
2. Now try it yourself

13

Lifecycle

1. Please watch the tutorial 2-Lifecycle(D:\Data\Haob\ Sujet)
2. Now try it yourself

Note: The four lifecycle phases here is the example You can create one by your self by importing the last lifecycle phases model and then rename it by voice input

14

Context

1. Please watch the tutorial 3-Context(D:\Data\Haob\Suje\)
2. Now try it yourself

15

Stakeholders and system

1. Please watch the tutorial 4-Stakeholder And System(D:\Data\Haob\Suje\)
2. Now try it yourself

Note: If you can't find the stakeholders you want in the menu during the later exercise, you can go down to the end of the stakeholder menu where you can find two primitives cube and sphere. You can customize their name to make it a substitute for the missing model.

16

Interface and function

1. Please watch the tutorial 5-InterfaceAndFunction (D:\Data\Haob\ Sujet)

2. Now try it yourself

Note: The moment you create an interface is the moment you determine the direction of the interface. The direction of the flow is always from the red port (out port) to the blue port (in port)

Example pour décrire les **functions** du telescope.

- pour l'interface entre le telescope et la source extérieure de puissance électrique: **Le telescope doit être alimenté par une énergie électrique**
- Pour l'interface entre le telescope et l'utilisateur: **Le telescope doit recevoir la référence de l'objet celeste à observer**

17



Exercise



Exercise with toothbrush system

Now please complete the tasks below:

1. Define the life cycle of the Connected Electric Toothbrush System
2. Define the external entities for a context of use (e.g., at home) of the Connected Electric Toothbrush System
 - Stakeholders
 - External systems
3. For a context of use (e.g., at home), define the external interfaces (ports and flows) and between the Connected Electric Toothbrush System and the external entities and functions of the Connected Electric Toothbrush system and the external entities.

19



Extra Training and Exercise

(If you have more than 20 mins left)

D'ont stop the Unity project yet otherwise you will lose all the work you have done during the exercise



Scenario

1. Please watch the tutorial 6-Scenario (D:\Data\Haob\ Sujet)

2. Based on your previous work for the Connected Electric Toothbrush System

For the context of use (e.g., at home), define the functional scenarios (functional chains) of the Connected Electric Toothbrush System.

- Scenario : To brush teeth

Note: You should always begin with red port then blue port as it is the correct direction of the interface, which also means you should always finish with blue port. So for one scenario, you should click by the order of the port color as red, blue, red, blue, red, blue.....red, blue

21



Questionnaire



Questionnaire

Now please fill out the questionnaire by using the link below to evaluate the interface

<https://questionpro.com/t/AYH7LZ0awP>



Appendix K Multi-Person Controlled Experiment Slides



Immersive Human-Computer Interface for Modelling Systems Architectures: Multi-Person Controlled Experimentation

Haobo Wang & Romain Pinqu  – Haobo.wang@grenoble-inp.fr



Motivations & Objectives

Motivations

- Powerfull MBSE modelling interfaces, but with a steep learning curve (abstract representations, limited modelling space, etc.)
- What could be the future MBSE modelling interfaces in 5 to 10 years?

Researchobjective

- Our research studies concentrate on new interactive interfaces to model the architecture of engineered systems naturally.
- To conduct exploratory and formative research to improve human-model interactions rather than discover "the scientific truth".

Multi-personcontrolledexperimentationobjective

- To compare the immersive modelling interface and 2D diagrammatic interface from the following perspectives:
 - User satisfaction
 - Visual notation
 - Cognitive load
 - Group dynamics

Multi-person controlled experimentation Protocol

1. Consent form (2')
2. Overview of a Wearable On-Body Drug Delivery System (5')
3. Introduction to the mission of a new Wearable On-Body Drug Delivery System (8')
4. List of system architecture tasks to achieve (5')
5. Exercise with Wearable On-Body Drug Delivery System in Capella 3D immersive interface (25')
6. Exercise with Wearable On-Body Drug Delivery System in 3D immersive interface / Capella 3 (5')
7. Post-experimentation questionnaire (20')

3

Overview of a Wearable On-Body Drug Delivery system

You will watch 2 introductory videos on wearable on-body drug delivery systems. While watching the video, try to identify some architectural elements of both systems, especially stakeholders, external systems, external interfaces, functions, etc.



BD Medical Pharmaceutical Systems: <https://www.youtube.com/watch?v=LhusP5bT3YU>



BD Medical Pharmaceutical Systems: <https://www.youtube.com/watch?v=oQNUuDI1vO4>

4

System Mission

What? The wearable on-body connected drug delivery system is an automated single-use injection device, filled at the hospital, to automatically administer 2-10 mL of a drug to cure chronic diseases and with frequent or higher autoimmune diseases and immuno-oncology drug viscosity within 2-10 minutes. It must also record injection data (device ID, injection start time, and injection end time) to document the correct completion of the drug delivery process, that is, detect deviation from the prescribed injection protocol, without modifying the user injection experience.

Why? The system is needed to replace an infusion pump, which performs drug infusions over hours or days and to allow patients to be active while receiving medication continuously, discreetly and comfortably.

Who? Patients with chronic autoimmune and immuno-oncology diseases.

Where? The drug delivery system is intended to be used at any place for basic activities of the daily life of a patient.

When? The drug injection will start 24 hours after the application of the device to the patient by a health care practitioner (oncologist, nurse, etc.) at the hospital.

5

Tasks

As a system architect, you will partially co-develop the architecture of the Wearable On-Body Drug Delivery System. The scope of the system analysis will be limited to the system level and pave the way to the definition of system requirements.

During the collaborative session, you must achieve 4 tasks:

1. Define the life cycle of the Drug Delivery System
2. Define the external entities for a context of use (e.g., at home) of the Drug Delivery System
 - External stakeholders
 - External systems
3. For a context of use (e.g., at home), define the external interfaces (ports and flows) between the Drug Delivery System and the external entities.
4. For the context of use (e.g., at home), define the functional scenario (aka. functional chains) of the Drug Delivery System.
 - Scenario : To inject a drug dose

6

Questionnaire

Link to the online questionnaire for Capella and 3D immersive interface
<https://questionpro.com/t/AYH7LZ0b7J>



Appendix L Formative Usability Tests Slides



Immersive Human-Computer Interface for Modelling Systems Architectures: Formative Usability Testing

Haobo Wang & Romain Piquié – romain.piquie@grenoble-inp.fr



Protocol

1. **Introduction(15')**
 - A. Research motivations and objectives (5')
 - B. Formative Usability Testing Protocol (5')
 - C. Consent form (5')
2. **Multipersonformativeusertesting(45')**
3. *Break (10')*
4. **Singlepersonformativeusertesting(60')**
5. *Break (10')*
6. **Post-formativeusertestingquestionnaire(15')**
7. **Post-formativeusertestingsemi-directedinterview(30')**

Motivations & Objectives

Motivations

- Lack of semiotic clarity due to numerous instances of symbol redundancy or overload.
- Poor discriminability of visual notations that consist of rectangle variants and connecting lines that are visually too similar
- Inordinate effort transposing models from rigorous tools to non-structured formats to overcome the acceptance problem
- Iconic graphics support significantly higher levels of understanding than highly codified symbolic notations

Research objective

- Our research studies concentrate on new interactive interfaces for modelling the architecture of engineered systems.
- To conduct exploratory and formative research to improve human-model interactions rather than discover “the scientific truth”.

Formative usability testing objective

- To evaluate how you perceive an immersive modelling interface as a Model-Based System Architecture expert.
- To clarify the requirements and design of the immersive modelling interface to satisfy your MBSA experts’ needs better.

3

Protocol

1. Consent form (5')
2. Multi-person formative user testing (45')
3. Break (10')
4. Single-person formative user testing (45')
5. Break (10')
6. Post-formative user testing questionnaire (15')
7. Post-formative user testing interview (45')

4

Consent Form

Autorisation de captation et de traitement des données collectées dans une recherche pour évaluer une interface de conception

Cher(e) participant(e),

Nous vous invitons à participer aux recherches menées par M. Haobo WANG à Grenoble INP-GSCOP UMR CNRS. Avant de décider de participer à cette étude, il est important que vous compreniez les objectifs de celle-ci, ce qui est attendu de votre part, et les implications de votre participation. Veuillez prendre un moment pour lire attentivement ce formulaire de consentement.

Objectif de l'étude :

L'objectif de cette étude est d'évaluer un prototype d'interface de modélisation d'une architecture système. Votre contribution en tant qu'expert en développement de système est essentielle pour l'évaluation du prototype.

Procédure :

Cette étude devrait durer approximativement 3 heures avec deux passes de 10 minutes. Au cours de cette étude, nous allons vous demander de réaliser des activités de modélisation d'architecture système pour deux systèmes différents avant de répondre à un questionnaire puis de participer à un entretien semi-dirigé.

Risques et avantages :

Il n'y a pas de risques importants associés à cette étude. Vous pourriez contribuer aux recherches réalisées dans la thèse de doctorat « Environnement virtuel immersif pour l'architecture de systèmes techniques ». Cependant, il n'y a pas de garantie de bénéfices personnels.

Confidentialité :

Si vous acceptez de participer à cette étude, vos discussions orales et votre activité de modélisation seront enregistrées pendant les deux exercices, ainsi que notre échange oral pendant l'entretien final. Toutes les données brutes recueillies seront considérées, les résultats anonymes dérivés de l'analyse des données brutes seront publiés dans le manuscrit de thèse de M. Haobo Wang et pourront être utilisés dans des articles scientifiques. Votre nom ne sera pas associé à vos réponses dans les résultats publiés. Vos données brutes seront conservées en toute sécurité et accessibles uniquement aux membres de l'équipe de recherche (Haobo Wang, Romain Pinquie et Frédéric Hechi).

Droits des participants :

Votre participation à cette étude est volontaire. Vous avez le droit de refuser de participer ou de retirer votre consentement à tout moment, sans aucune conséquence négative. Si vous avez des questions ou des préoccupations concernant cette étude, veuillez contacter Haobo WANG par haobo.wang@grenoble-inp.fr

Consentement :

Je, soussigné(e) Nom : _____ Prénom : _____

ai lu et compris l'information sur la recherche en intégralité. J'ai obtenu des réponses à toutes les questions que j'ai posées et je suis d'accord pour apporter ma contribution à ce projet dans les limites des conditions acceptées ci-dessous.

- Je reconnais consentir à la captation, l'enregistrement et le traitement des données (vidéo/image) dans le cadre de ma participation volontaire à la recherche concernant la thèse « Environnement virtuel immersif pour l'architecture de systèmes techniques » de M. Haobo Wang sous la responsabilité scientifique de M. Romain Pinquie et pour une durée de 3 heures.
- J'autorise la diffusion des résultats au cours de colloques et de formations sans mise à disposition à un public.
- J'autorise la mise à disposition des résultats à la communauté scientifique pour mener d'autres recherches.
- J'autorise la diffusion des résultats dans des publications scientifiques avec accès à un public.

Toute reproduction ou utilisation de ces informations ne devra pas, y compris par leur légende ou leur contexte de présentation, porter atteinte à ma réputation, à mon honneur ou à ma dignité.

Conformément au Règlement général sur la protection des données (RGPD), je dispose d'un droit d'accès aux données me concernant et de cas échéant de suppression de ces données, que je peux exercer auprès de M. Romain Pinquie. (*) J'ai été informé(e) que le droit de suppression des données ne concernant ne peut pas s'exercer si cela porte préjudice aux objectifs poursuivis, notamment modifier le protocole ou les résultats de la recherche.

Fait en deux exemplaires à _____ le _____

Signature

Un exemplaire de ce document vous est remis, un autre exemplaire est conservé et archivé sous la responsabilité du responsable scientifique.

5



Immersive Human-Computer Interface for Modelling Systems Architectures: Multi-Person Formative Usability Testing

Haobo Wang & Romain Pinquie – romain.pinquie@grenoble-inp.fr



Objective

To evaluate the use of the immersive modelling interface in a situation of collaborative architecture design.

7

Protocol

1. **Introduction(15')**
2. **Multipersonformativeusertesting(45')**
 - A. Overview of a Wearable On-Body Drug Delivery System. (5')
 - B. Introduction to the mission of a new Wearable On-Body Drug Delivery System. (5')
 - C. List of system architecture tasks to achieve. (5')
 - D. Exercise (25')
 - E. Give you a quick overview of further modelling capabilities (new context, requirement, subsystems (5')
3. *Break (10')*
4. **Singlepersonformativeusertesting(60')**
5. *Break (10')*
6. **Post-formativeusertestingquestionnaire(15')**
7. **Post-formativeusertestingsemi-directedinterview(30')**

8

Overview of a Wearable On-Body Drug Delivery system

You will watch 2 introductory videos on wearable on-body drug delivery systems. While watching the video, try to identify some architectural elements of both systems, especially stakeholders, external systems, external interfaces, functions, etc.



BD Medical Pharmaceutical Systems: <https://www.youtube.com/watch?v=LhusP5bT3YU>



BD Medical Pharmaceutical Systems: <https://www.youtube.com/watch?v=oQNUuDI1vO4>

9

System Mission

What? The wearable on-body connected drug delivery system is an automated single-use injection device, filled at the hospital, to automatically administer 2-10 mL of a drug to cure chronic diseases and with frequent or higher autoimmune diseases and immuno-oncology drug viscosity within 2-10 minutes. It must also record injection data (device ID, injection start time, and injection end time) to document the correct completion of the drug delivery process, that is, detect deviation from the prescribed injection protocol, without modifying the user injection experience.

Why? The system is needed to replace an infusion pump, which performs drug infusions over hours or days and to allow patients to be active while receiving medication continuously, discreetly and comfortably.

Who? Patients with chronic autoimmune and immuno-oncology diseases.

Where? The drug delivery system is intended to be used at any place for basic activities of the daily life of a patient.

When? The drug injection will start 24 hours after the application of the device to the patient by a health care practitioner (oncologist, nurse, etc.) at the hospital.

10

Tasks

As a system architect, you will partially co-develop the architecture of the Wearable On-Body Drug Delivery System. The scope of the system analysis will be limited to the system level and pave the way to the definition of system requirements.

During the collaborative session, you must achieve 4 tasks:

1. Define the life cycle of the Drug Delivery System
2. Define the external entities for a context of use (e.g., at home) of the Drug Delivery System
 - External stakeholders
 - External systems
3. For a context of use (e.g., at home), define the external interfaces (ports and flows) between the Drug Delivery System and the external entities.
4. For the context of use (e.g., at home), define two functional scenarios (aka. functional chains) of the Drug Delivery System.
 - Scenario A: To inject a drug dose
 - Scenario B: To monitor the injection of a drug dose (if not running out of time)

11



Immersive Human-Computer Interface for Modelling Systems Architectures: Single-Person Formative Usability Testing

Haobo Wang & Romain Piquié – [✉ romain.piquie@grenoble-inp.fr](mailto:romain.piquie@grenoble-inp.fr)



Objective

To evaluate the use of the immersive modelling interface in a situation of individual architecture design.

13

Protocol

1. **Introduction(15')**
2. **Multipersonformativeusertesting(45')**
3. *Break (10')*
4. **Singlepersonformativeusertesting(60')**
 - A. Introduction to the mission of a new Connected Electric Toothbrush System. (3')
 - B. List of system architecture tasks to achieve. (2')
 - C. Training (20')
 - D. Exercise (35')
5. *Break (10')*
6. **Post-formativeusertestingquestionnaire(15')**
7. **Post-formativeusertestingsemi-directedinterview(30')**

14

System Mission

What? The electric toothbrush system has the mission is to clean users teeth. The electric toothbrush will also monitor brushing data so as to increase the brushing efficiency by sending recommendations to the user through a mobile application. The electric toothbrush system will be connected to collect the following data: Start time; End time; % of time used for each mode; and Time spent with a brushing pressure exceeding the threshold specified by healthcare recommendations.

Why? To rapidly (approximately between 2 and 3 minutes) clean users teeth.

Who? Teenagers and adults from 14 to 99 years old.

Where? At any place (home, airport, train station, hotel, etc.) but mainly in the bathroom.

When? Three times per day (after breakfast, lunch and dinner).



15

Tasks

As a system architect, you will partially develop the architecture of a Connected Electric Toothbrush System. The scope of the system analysis will be limited to the system level and pave the way to the definition of system requirements.

During the individual session, you must achieve 4 tasks:

1. Define the life cycle of the Connected Electric Toothbrush System
2. Define the external entities for a context of use (e.g., at home) of the Connected Electric Toothbrush System
 - Stakeholders
 - External systems
3. For a context of use (e.g., at home), define the external interfaces (ports and flows) between the Connected Electric Toothbrush System and the external entities.
4. For the context of use (e.g., at home), define two functional scenarios (functional chains) of the Connected Electric Toothbrush System.
 - Scenario A: To brush teeth
 - Scenario B: To monitor the brushing activity (if not running out of time)

16



Immersive Human-Computer Interface for Modelling Systems Architectures: Semi-Directed Interview

Haobo Wang & Romain Piquié – romain.piquie@grenoble-inp.fr



Post-Formative Usability Testing Interview

- What do you think about the use of system architecture modelling software and languages?
- Do you think that current system architecture models are well received by notational nonexperts who are unfamiliar with such graphical notations?

- How did you feel about the use of the immersive modelling interface?
- What do you think about 3D representations compared with diagrammatic representations?
- How did the immersive modelling interface helped you to complete your tasks?
- What barriers did you face in attempting to use the immersive modelling interface to complete your tasks?
- Do you think that the immersive modelling interface facilitates a shared understanding?
- What did you think about the interactive devices (HMD and Powerwall)?

- Could you imagine using a similar modelling interface for system architecture in the future?
- In which situations would you want to use such an application?

- If you were to integrate such an immersive modelling interface in your work, how would you improve it?

19