HELSINGIN YLIOPISTO

Instructions on Small Screens

Analysing the Multimodality of Technical Communication Through a Design Experiment

Petri Ahola
Master's Thesis
English Philology
Department of Languages
University of Helsinki
September 2020

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistinen tiedekunta / Kielten osasto		
Tekijä – Författare – Author AHOLA, Petri Olavi		
Työn nimi – Arbetets titel – Title Instructions on Small Screens: Analysing the Multimodality of Technical Communication Through a Design Experiment		
Oppiaine – Läroämne – Subject Englantilainen filologia		
Työn laji – Arbetets art – Level Maisterintutkielma	Aika – Datum – Month and year syyskuu 2020	Sivumäärä– Sidoantal – Number of pages 97

Tiivistelmä – Referat – Abstract

Tässä tutkielmassa analysoin teknisen viestinnän multimodaalisuutta kokeellisen suunnittelun avulla. Kokeessani suunnittelen ja konvertoin älylasien pienelle näytölle kolme lyhyttä KONE Oyj:n asennus- ja huolto-ohjetta. Vaikka käytän kokeessani älylaseja, tutkimuksen näyttö voisi periaatteessa olla mikä tahansa pieni näyttö, esimerkiksi älypuhelin tai älykello, jonka ajantasainen sisältö on teoriassa helpommin kuljetettavissa mukana kuin paperille tulostettu perinteinen PDF-ohje. Konvertoin ohjeet käyttäen kahta teoriaa: visuaaliset ohjeet (Gattullo et al. 2019) ja minimalismiheuristiikka (van der Meij ja Carroll, 1998). Ymmärtääkseni paremmin ohjeiden koko käyttökontekstia, rakennan konversioiden testaamiseen yhteistyönä KONE Oyj:ssä käyttäjätestiympäristön simuloimaan ammattimaista hissin asennus- ja huoltoympäristöä.

Vaikka nykytekniikka mahdollistaa digitaalisten, pienten näyttöjen käytön, ohjeiden tarkoitus ei muutu: niiden pitää ymmärrettävästi auttaa lukijaa suorittamaan tehtävänsä. Täten konversio- ja suunnitteluteorioiden vastapainoksi multimodaalisuuden teoriat (esimerkiksi, Bateman, Wildfeuer ja Hiippala, 2017) auttavat analysoimaan konversioiden ymmärrettävyyden eroja systemaattisesti. Käytän tutkielmassani multimodaalisuuden teorioita ymmärtääkseni konversioiden vaikutukset ohjeiden ymmärrettävyyteen. Multimodaalisuuden teorioiden avulla tunnistan ohjeiden käyttötilanteen, käytetyn median (älylasit) ominaisuudet, sekä rajaan varsinaiseksi tutkimuskohteekseni konvertoiduilta ohjenäytöiltä tunnistamani semioottiset moodit ja niiden vaikutukset konvertoitujen ohjeiden ymmärtämiseen.

Johtopäätöksinä esitän, ettei yksittäisiä konvertoituja ohjenäyttöjä tutkimalla synny mimimalismiheuristiikan osalta ymmärrettävyyden kannalta merkittäviä eroja lähtötilanteen PDF-ohjeeseen nähden, lukuun ottamatta muutamien helposti pääteltävien kohtien poisjättämistä. Yleisesti ottaen molemmissa konversioissa älylaseille siirtyy multimodaalisesti samankaltainen, kaksiulotteista sivunäkymää hyödyntävä ohje kuin lähtötilanteen PDF. Koska toinen tutkimani teoria, visuaaliset ohjeet, perustuvat verbien korvaamiseen symboleilla, symbolien ymmärrettävyys korostuu merkittävänä erona visuaalisten ohjeiden käytettävyydessä. Johtopäätöksiä selventää, etten hyödynnä älylasien kaikkia ilmaisukeinoja, kuten liikkuvaa kuvaa ja ääntä, koska kokeessani huomioin kustannustehokkaan, teollisten ohjeiden tuotantoprosessin. Lopuksi ehdotan erityisesti teknisen viestinnän viitekehyksessä jatkotutkimuksen aiheiksi uusien digitaalisten medioiden kaikkien ominaisuuksien ja niiden multimodaalisten käyttötilanteiden tutkimista ja hyödyntämistä, pienien näyttöjen sisällöntuotannon standardisoinnin tutkimista ja kehitystä, sekä symbolien ymmärrettävyyden tutkimista.

Avainsanat – Nyckelord – Keywords
multimodaalisuus, tekninen viestintä, digitaaliset ohjeet, minimalismiheuristiikka, visuaaliset ohjeet

Säilytyspaikka – Förvaringställe – Where deposited
Keskustakampuksen kirjasto

Muita tietoja – Övriga uppgifter – Additional information
Tutkielman nimi suomennettuna
Ohjeet pienillä näytöillä: teknisen viestinnän multimodaalisuuden analysointia kokeellisen suunnittelun avulla

Table of contents

List of tables

List of figures

Acknowledgements

In addition to Tuomo Hiippala's thoughtful and inspiring supervision of my thesis, I would like to thank my manager, Teemu Ruotsalainen, for opening the door to combine my daily work at KONE Oyj and my master's thesis for the benefit of both. Equally, I thank my supervisor at KONE, Sanni Siltanen, and my colleague, Hanna Heinonen, for giving me the opportunity to join research forces, providing me invaluable motivation, guidance and access to the world of augmented reality equipment. Thank you to Gokul and Jayaraj for supporting with symbol creation in my design experiment. All installation and maintenance instructions displayed in this thesis, including text, images and screenshots, are property of KONE Oyj and used by permission.

1 Introduction

> "As a controlled language, Newspeak limits the user's communications (thought, spoken, and written) with a vocabulary that diminishes the intellectual range allowed by Oldspeak (Standard English)" (Orwell, 1987).

Traditional paper manuals or laptops with large screens are not always practical when performing a professional installation or maintenance task. The physical working location can limit the use of a laptop, and with paper, the print taken some time ago might be out-of-date or relevant parts or references can be missing. This is especially true with traditional mechanical industry, such as aviation or elevator maintenance and installation. For example, there is a risk of falling in a narrow and sometimes poorly lit elevator shaft, so using a laptop or a large tablet can even count as a safety risk. A simple task of entering an elevator shaft is considered dangerous, so moving back and forth to get missing papers must be eliminated.

Kaasinen et al. (2018, p. 11) describe the challenges that technicians face in their work environment. The work sites, tools, reporting, communication and the maintained equipment itself are increasingly diverse. The site environment can be noisy, hot, greasy, humid and contain high and poorly lit places.

Figure 1 Technician reading instructions on smart glasses (simulated test experiment scenario)

With the age of mobile devices that we all carry with us, the information medium and publication channel is easily available, but the available space for the content on their small screens is typically small, at least smaller than A4 paper. In addition to mobile phones, powerful and reasonably priced smart glasses are available (Eschen et al., 2018, p. 156). Despite rapid technological advances with new display medias, content (the instructions) is still often designed and produced for paper or for larger screens. Regarding the instructions of the technicians, Kaasinen et al. (2018, p. 11) state that the instructions are typically in PDF format, without any redesign for small, digital screens. They continue to point out that a paper print does not update itself with new revision information, so the technician must constantly check that the ever-changing equipment at the work site and the printed instructions match. At the site, it is too late, because of limited possibilities for obtaining updated instructions. Thus, there is a growing need for guidelines on how technical communicators should work with the new technological devices, but few exist (Tham et al., 2018, p. 179). In fact, authoring of technical content for new devices even counts as a problem. Technical writers lack guidelines to follow (Gattullo et al., 2017, p. 1304).

The practical problems include, for example, how to make the instructions readable, skimmable and understandable in the limited space offered by small screens. Studies have pointed out that scrolling should be avoided, and full screen and sensory modalities should be utilised (Churchill and Hedberg, 2008, p. 892). Thus, the use of alternative forms of expression compared to written, linear text in PDFs must be considered, for example, the use of symbols, images, diagrams, information graphics, videos, sounds, voice and haptic technology. Yet, it is important to note the difference between sensory modalities, for example, sight, hearing or touch, and semiotic modes that are the actual meaning-making tools in multimodality. Despite the advances in using sensory modalities in, for example, smart glasses, the goal of an instruction is to get the instructed message across to the users, so that they can accomplish their goals. Therefore, the meaning-making aspect of the semiotic modes and multimodality is as important as ever, and an aspect that should not be forgotten when working with the sensory modalities that the current augmented reality (AR) devices are technically capable of supporting. Yet, it is important to understand how the new small screen devices participate in the semiotic meaning-making process, as the available modes of communication depend on the underlying medium (the small screen device) and its material properties. For example, the capabilities of particular smart glasses and the software installed on them can limit or enable their possibilities for utilising various semiotic modes, and sensory modes.

The challenges described above are related to the digital revolution concept called Industry 4.0. Industry 4.0 is an established concept within the discussion of industrial revolution and its research. Industry 4.0 is conceptualised as the next evolution of the industrial revolution continuum that started at the end of the 18th century with the invent of water- and steam-powered factories. Industry 4.0 is the result of the evolution of information technology, personal computers, networks, cloud computing, virtual cyberspaces, and the Internet of things and services that will impact and force industry to change and become Industry 4.0 (Kagermann et al., 2013, p. 15). Technical communication is also included in Industry 4.0, as instructions need to be digitalised and made more visual than before, possibly utilising also other sensory modalities than just visual. The new type of digital and visual instructions are part of Industry 4.0 technologies, enabling accessing, storing and displaying technical instructions (Gattullo et al., 2017, p. 1304). For example, in Industry 4.0 maintenance, technicians use network integrated platforms to perform remote maintenance with their mobile devices (Kagermann et al., 2013, p. 67). Here, digitally integrated instructions enable communication between machines and humans. In case of an error, there must be case-specific, exact instructions immediately available, without having to search for them from a traditional manual (Gattullo et al., 2019, p. 276). Technical communication has already responded to the term Industry 4.0 by coining Information 4.0, where smart content responds to Industry 4.0 users' needs, and is one of the most important products of Industry 4.0 (Gallon and McDonald 2016, p. 15).

Technical communication itself is a broad field that includes any technical, technological or instructive form of communication for users to achieve their goals (Society for Technical Communication, 2019). Traditionally, technical communication has been designed to be published on A4 sized paper, with relatively large amount of space available for content. Even with content published on computer screens and as online helps embedded in software, there has always been space available for the content.

Technical communication as a profession focuses on clear, short and to the point content, often aided by a company style guide as the writing guideline. This naturally lends itself well to small screens, yet the focus has been on A4 size PDF content creation. Several industry standards exist, on which company style guides have been traditionally based on. For example, Controlled Natural Languages (CNL), such as ASD Simplified Technical English (STE), Caterpillar Technical English and IBM Easy English, as well as Microsoft Writing Style Guide and Chicago Manual of Style are among the most well-known ones. From the small screen point of view, perhaps one of the most widely known theories that technical communicators use is minimalism. With the basic idea that users only need directly useful, action-oriented, but not comprehensive, information to accomplish tasks (Lodor, 1999, p. 49), the principles of minimalism sound like a good starting point when producing instructive content for small screens.

To understand designing technical communication content for the challenging context stated above, the scope of this thesis is to evaluate two technical communication theories for Industry 4.0 instruction production, specifically to overcome the obstacles of small screen instructive content design. In addition to evaluating the design aspects of the theories, this thesis also sets out to understand the *effects* of such small screen designs to the *users* of the instructions by using theories of multimodality. The evaluated theories are the following:

- Visual manual by Gattullo et al. (2019)
- Minimalism heuristics by van der Meij and Carroll (1998)

Visual manual is a theory to convert traditional, paper-oriented instructions to augmented reality -oriented instructions that are usable in the Industry 4.0 context (Gattullo et al., 2017; Scurati et al., 2018; Gattullo et al. 2019). The theory promotes replacing existing text with graphics, symbols and icons (Gattullo et al., 2019, p. 277).

Minimalism heuristics convey the design principles presented by van der Meij and Carroll (1998) for writing minimalistic instructions. According to the minimalistic approach, it is not the full system that gets instructed, but a user-centred approach to only instruct what the users need to accomplish the specific tasks they have in mind. Additionally, error prevention information is added to further aid the user to successfully accomplish the task. Since unnecessary descriptive text is omitted, minimalism should lend itself well for small screen. Additionally, minimalism heuristics complement the modern visual manual theory by providing a more researched counterpart.

This thesis combines experimental and qualitative methods by having authored versions of task instructions (the data), whose design is theoretically-informed. To achieve this, I converted three originally A4 paper format instructions, two installation tasks and one maintenance task, as test material for testing the visual manual and minimalism heuristics theories. I published the converted instructions, in addition to a conventional version with no redesign, to smart glasses for multimodal analysis and discussion. This design experiment is part of a larger research initiative inside KONE Oyj. The design experiment in this thesis serves as a foundation for an extensive user test, prepared and performed together with my KONE colleagues, Sanni Siltanen and Hanna Heinonen. The quantitative results of the user test falls outside the scope of this thesis, but the practical arrangements and selected, qualitative observations are included. They add the *context* for analysing the overall *situation* where readers use the instructions and provide further understanding how the *situation* affects understanding the instructions. Although at the centre of the small screen user experience, this thesis also excludes how to search and find the relevant information content to be displayed on the small screen. Also excluded are all technical means for producing the content, and how to technically export and display the content on the screen. User interface design and evaluating different small screen devices themselves are also excluded.

To get a theoretical understanding of the full context of this thesis, Chapter 2 introduces the technical communication industry and its concepts related to this thesis. Chapter 2 also describes the selected main design theories, visual manual and minimalism heuristics. Theories of multimodality tie Chapter 2 together by providing a synthesis for the other discussed theories, and by providing analytic tools for understanding the effects of my design experiment to the users of the instructions. Chapter 3 describes the design experiment and specifies the used methods and materials in detail. To continue with the design experiment, Chapter 4 takes us into the action of conducting the design experiment according to visual manual and minimalism heuristics by describing the individual conversion actions and their reasons, and analyses the results with the theories of multimodality. The discussion in Chapter 5 widens the focus by discussing the authoring and test user experiences in the context of professional technical communication and multimodality, and points out avenues for further research. Finally, the conclusion in Chapter 6 summarises the main points of the design experiment and analysis.

2 Theoretical perspectives on instructions on small screens

This chapter lays the theoretical framework where the studied design theories sit in, and how their effects can be analysed systematically with theories of multimodality. Section 2.1 gives a brief history of technical communication, highlighting the technical communication theories, guidelines and standards that are relevant to this thesis, such as Darwin Information Typing Architecture (DITA), Simplified Technical English (STE), Information Mapping and minimalism for technical communicators. This chapter then continues to open up the visual manual theory (see Section 2.2) and minimalism heuristics (see Section 2.3). These are the theories utilised in my design experiment described in Chapter 3 and Chapter 4 for converting paper format instructions into the small screen of a smart glass. The theories of multimodal communication are presented in Section 2.4. They frame the theoretical synthesis of this thesis by providing tools, concepts and terminology for structurally understanding how converting the instructions according to the visual manual theory and minimalism heuristics affect *the users* of the instructions from the semiotic, meaning-making point of view.

2.1 Technical communication

Technical communication is a broad field that includes any technical, technological or instructive form of communication to users for enabling them to achieve their goals (see, for example, Jayaprakash, 2008). Technical communicators, also called as technical writers or technical authors, study, plan, design and create information, documents, or content of any form about technical or specialised processes, products or services. The produced information is targeted to a specific audience, for example, a user manual typically instructs a consumer end user how to use a product. A repair manual on the other hand, instructs a professional maintenance technician how to repair a broken product. The resulting information product can take several forms, ranging from classic A4 paper or book format to modern extended reality (XR) outputs (Society for Technical Communication, 2019).

The origin and early historical events of technical communication are difficult to pinpoint. The profession has been recognised and called technical writing (nowadays technical communication) only since World War II, with the evolvement in electronics, motors, engineering, and medicine. Eventually with computers, technical communication requirements and skills started to advance. Yet, the elements of technical communication are reflected already in historical and ancient scripts. Some examples of early instructional books on technical communication are considered to be T.A. Rickard's *A Guide to Technical Writing* (1908) and Samuel Chandler Earle's *The Theory and Practice of Technical Writing* (1911). Even though many colleges in the US offered technical communication as degree programs already in the 1960s, the U.S. Department of Justice ruled technical communication to be a profession only in 1980s. In the mid-1990s, ISO 9000 certification required companies to have documentation on their products, thus increasing the work opportunities for technical communicators (Jayaprakash, 2008, pp. 17–18).

As for the end product of the profession, a good instruction is complete, easy-to-read and understand, and easy to access and locate the required information. There are techniques, guidelines, and processes for writing good instructions, but eventually the readers only care about the content that answers their questions and gets their job done. Thus, it is not good enough that instructions are grammatically and technically correct, they must also serve the readers' need as described above. To create such user-friendly instructions requires background work and thought. Additionally, different users and tasks require different content. Advanced users only prefer to see hard-to-remember details, whereas novice users require more generic content to instruct performing the basic task (Jayaprakash, 2008, pp. 230–233).

For authoring instructions, technical communication theories, guidelines and standards, such as minimalism, Simplified Technical English (STE) and Darwin Information Typing Architecture (DITA) XML have been researched and developed. Today, as introduced already, the established ways of authoring technical communication are under pressure from moving away from the page-based media. For example, the visual manual theory studied in this thesis does not belong to the "classical canon" of technical communication theories, but it is a more modern take, geared especially for smart glasses. Juxtaposing the classical minimalism heuristics and the modern visual manual theory for small screen design creates a setting where one can expect to see differing results.

To help understanding the background, the following subsections briefly introduce a selection of the "classical" theories and standards of technical communication that the main design experiment theories of this thesis, visual manual and minimalism heuristics, refer to. After understanding the background, Sections 2.2 and 2.3 then describe the main design experiment theories of this thesis in detail.

2.1.1 Modular design and DITA XML

Darwin Information Typing Architecture (DITA) is an extensible markup language (XML) -based standard for modular writing. DITA defines modules called topics, which are long enough articles to make sense on their own, but short enough to cover only one topic or procedure. A collection of topics is typically published as a traditional PDF manual or a collection of HTML pages (Bellamy, Carey and Schlotfeldt, 2012, p. 25). To cover the most common user needs, DITA has three main topic types. A **task** topic type describes procedural instruction, **concept** topic type provides background or conceptual information, and **reference** topic type lists quick reference information. The division to topic types helps authoring and reading the content, because users typically do not need these topic types bundled together, for example, a procedure buried in a reference information table (Bellamy, Carey and Schlotfeldt, 2012, p. 28). Within each topic type, the available XML elements and their order in the topic are predefined, enabling consistency between different topics and authors.

DITA is directly linked to the theories evaluated in this thesis, minimalism heuristics and visual manual. Bellamy, Carey and Schlotfeldt (2012, p. 32) state that if authors follow minimalistic writing principles, the DITA topics written are more effective and contain only the information that the users need, without anything extra. Gattullo et al. (2019, p. 281) note that DITA is a well-established technological solution to structure, author and publish content. For Industry 4.0, DITA as structured writing enables single-source publishing, where the same single-source XML module that gets published to a PDF manual can also be published as an HTML page, or to a small screen medium, mitigating the need to rewrite the same content separately for different output media.

2.1.2 Controlled languages and ASD Simplified Technical English

From technical communication point of view, a controlled language (CL) has been designed to make a language easily understandable and translatable by non-native readers. To achieve this, a CL uses a restricted version of a natural language, with a defined subset of grammar and lexicon, and adding the terminology of the technical communication domain in question. CLs have a background in language teaching, dating from 1930s. The best-known controlled language is AECMA Simplified English, currently known as ASD Simplified Technical English (STE). ASD STE was originally designed for aerospace industry, but developed and adopted by several other companies as well (Kittredge, 2005, p. 11).

Kuhn (2014, pp. 133–139) lists and compares several English-based CLs and introduces twelve as the most influential and well documented ones, for example, Basic English, Caterpillar Fundamental English, FAA Air Traffic Control Phraseology, and ASD STE. Due to its established status among CLs, ASD STE was chosen to be included in the visual manual theory by Gattullo et al. (2019, pp. 278–279). I will return to the visual manual theory in Section 2.2.

2.1.3 Information Mapping

Information Mapping is a set of principles for analysing, organising and presenting information systematically. Information Mapping is sometimes mistakenly thought as a competitor of DITA XML, but they actually supplement each other. Even though both are modular information design frameworks, DITA focuses more on the technology solution (standard) and Information Mapping on information design. Therefore, they both can be adopted in the same technical communication system (Information Mapping, 2020).

Information Mapping features replacing traditional text paragraphs with information blocks. Transforming text paragraphs into information blocks is called chunking, which the visual manual theory by Gattullo et al. (2019, pp. 282–283) utilises further (see Section 2.2.2). Information blocks are organized as collections that are called

Information Maps. Visually, an information block has a title at the top, graphics at the centre. and text on the sides of the graphic. The hierarchical organisation of information blocks provide a table of context to present the context of the procedure (Gattullo et al., 2019, p. 281).

2.1.4 Minimalism for technical communicators

Minimalism in instructions stems from the era of time when computer usage started to become more widespread in the 1980s. John Carroll's *The Nurnberg Funnel* (1990) is considered as a seminal work that established minimalist design principles for instructions, and challenged existing theories to design and deliver instructional computer documentation (Lodor, 1999, p. 48). Lodor (1999, p. 48) states that instead of making users to learn all features of the system inside out before taking action, Carroll pointed out that computer users require action-oriented instructions that enable them to start working immediately. After *The Nurnberg Funnel* (Carroll 1990), Carroll and co-authors continued to investigate new directions and developments of minimalism in *Minimalism Beyond the Nurnberg Funnel* (Carroll 1998). Even though much has been written about minimalism, there has been and still is a need to develop *practical* minimalist style guides for design purposes to anchor minimalist principles into the technical communication process (Virtaluoto, Suojanen and Isohella, 2021, p. 28). As a remedy, van der Meij and Carroll (1998) and as the latest, Virtaluoto, Suojanen and Isohella (2021) have coined principles and heuristics for designing and analysing minimalistic instructions. Because of their established status, the design experiment of this thesis uses the original heuristics by van der Meij and Carroll (1998) and introduces them in Section 2.3.

The core idea of minimalistic writing is that the users of the instructions only need directly useful, action-oriented, but not comprehensive, information to accomplish tasks (Lodor, 1999, p. 49). Despite all modern digital development, it is the user-centeredness, in other words, knowing what the users need and what they do not need, that is the corner stone of minimalism in instructions (Virtaluoto, Suojanen and Isohella, 2021, p. 28). Regarding the modern Industry 4.0 instructions, minimalism still has much to offer (Virtaluoto, Suojanen and Isohella, 2021, p. 2).

2.2 Visual manual theory for small screen

After gaining a basic understanding of technical communication (see Section 2.1), let us have a detailed look on the first conversion theory utilised in the design experiment of this thesis: the visual manual theory by Gattullo et al. (2019). Gattullo et al. (2019, p. 277) indicate that although modern research considers that the visualisation of instructions for augmented reality (AR) purposes is aligned with Industry 4.0, the research lacks specific guidelines for converting existing paper documentation to visual manuals. Gattullo et al. (2019, p. 277) state that the existing methodologies (see, for example, Knopfle et al., 2005, Stock et al., 2005, Engelke, 2013, Gimeno et al., 2013, Erkoyuncu et al., 2017) focus on creating new content, thus lacking guidelines for reusing *existing* documentation. According to Gattullo et al. (2019, p. 277), the result is that the supportive information present in existing paper documentation, which helps interpreting the instruction, is not delivered to technicians. Thus, Gattullo et al. (2019) propose a guideline to convert (redesign) existing paper documents (text, printed or PDF format) to AR visual manuals as required by Industry 4.0. Visual manuals will have less text and more graphics and symbols. They also suggest that their guideline can be used by authors to write new AR instructions from scratch. Gattullo et al. (2017, p. 1304) note that their aim for text reduction equals reduced document update times, faster, more comprehensive and more intuitive reading experience for technicians, and easier translation, resulting in cost reduction for companies.

Gattullo et al. (2017, p. 1305) show the current evolution from a paper manual to a visual, augmented reality manual. They claim that the current industry process is only an *augmented text approach*. An augmented text approach includes migrating existing text from a paper manual to a digital manual, but without any redesigning, illustrating or rewriting. They propose that the first evolution should be *augmented images approach*. The augmented images approach requires redesigning and rewriting the documentation so that the authoring starts with visual content (images or videos), not with text. Only a minimum amount of text should be added where required. They claim that the goal of the evolution is *augmented reality approach*, where everything, design, images and text, start from scratch. Here, the content creation starts from placing the visual content, for example, 3D animations, CAD models and 2D graphic signs on a real visual AR scene. However, they point out that the augmented reality approach is not yet directly implementable due to lack of authoring standards to follow.

To remedy the lack of existing guidelines, Gattullo et al. (2019, p. 278) propose the visual manual theory as illustrated by Figure 2:

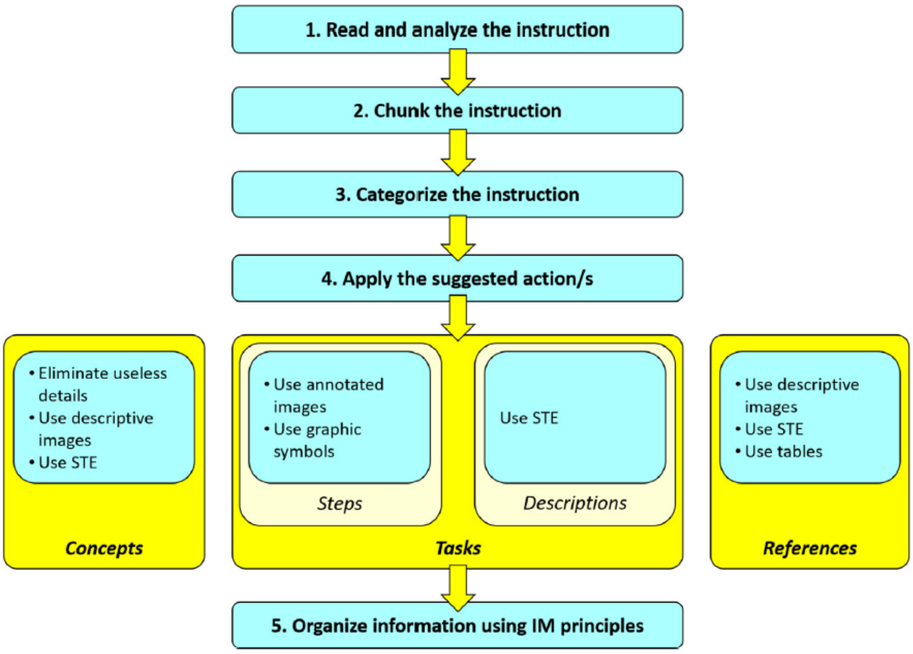


Figure 2: Converting existing instructions to visual manuals (STE = Simplified Technical English, IM = Information Mapping) (Gattullo et al., 2019, p. 278)

The visual manual theory also includes minimalistic ideas, as seen in Figure 2 ("Eliminate useless details"). In a visual manual, only the required information must be provided, and useless details can be left out. The use of too much text occludes the real scene (Gattullo et al., 2019, p. 282). With AR manual visualisation requirements in mind, Radkowski (2015, p. 342) suggests that task complexity is the determining factor. Complex instructions require more complex visual features, whereas simpler tasks can use only text.

To understand how the design experiment of this thesis (see Chapter 4) applies the visual manual main phases (see Figure 2) in practice, the following subsections explain the visual manual conversion phases step-by-step.

2.2.1 Phase 1: read and analyse the instruction

Phase 1 includes reading the existing instruction. The goal is to find suitable locations in the instruction for the chunking of phase 2 (Gattullo et al., 2019, p. 278). Finding locations for chunking is similar than the basic idea of modular writing, that is, dividing content into information modules. In DITA XML, content is divided and structured modularly into concept, task, and reference topics. Considering the scope of this thesis, a similar tool for *multimodal analysis* is Bateman's Genre and Multimodality model (GeM) (see Bateman, 2008, pp. 15–19, 25–26, 107–129) that decomposes documents and pages to units and defines layers for multimodal description. I will return to multimodality in instructions in Section 2.4.5.

In their case studies, Gattullo et al. (2019, pp. 281–282) explain phase 1 with an example instruction taken from a hydraulic valve pressure release instruction for elevators:

> “push the manual descent (MM) button to unload the pressure.”

They apply minimalistic theories when they delete the latter part of the step “to unload the pressure”. They assume that the user is aware of the target of the step, so it can be deleted from consuming space and causing occlusion in an AR view.

In their second case study, Gattullo et al. (2019, p. 282) show that phase 1 and phase 2 are actually intermittent. The goal of the reading and analysing of phase 1 is to isolate content into task, concept and reference topics according to DITA XML modularisation principles. This is the same as the chunking in phase 2.

2.2.2 Phase 2: chunk the instruction

Gattullo et al. (2019, p. 278) give the following simplified example for chunking:

> “1) ‘Turn the accumulator’, 2) ‘pour in oil to fill the cavity with the holes’; 3) ‘check that the membrane has no porosity’; 4) ‘If there is no leakage, it means that the accumulator has been charged correctly and is ready to be fitted on’”

In the example above, the content that gets chunked according to Information Mapping principles are single steps of a task procedure, but chunking can also mean modularising bigger entities of content into topics according to DITA XML, where each DITA topic is a chunk and can contain multiple steps (Gattullo et al., 2019, pp. 282–283).

2.2.3 Phase 3: categorise the instruction

For categorising the chunked instruction, Gattullo et al. (2019, pp. 282–283) use the DITA XML information type division to concept, task, and reference topic types. The categorisation helps to decide which conversion action to choose in phase 4, resulting in a different form in the output of the content.

The visual manual theory does not apply DITA XML in the strictest sense, but *adapts* it for categorising information. For example, the example steps (1), (2) and (3) quoted in Section 2.2.2 are categorised as the steps of a single task topic type, and step (4) as a conceptual description of the result of the task. In the hydraulic valve instruction quoted in Section 2.2.1, the single example step 1 of a task topic is categorised as task information type (Gattullo et al., 2019, p. 282). In contrast, Bellamy, Carey and Schlotfeldt (2012, p. 28) state that the whole topic, including all steps and their results, is categorised as a task, concept or reference topic type in DITA XML. Single steps or the results of single steps, or the results of the whole topic, are not categorised as tasks, concepts and references by themselves. Inside task topics, Gattullo et al. (2019, p. 278) further distinguish content into steps and descriptions.

2.2.4 Phase 4: apply the suggested actions

The suggested actions according to Gattullo et al. (2019, p. 278) are:

- Optimize text with ASD Simplified Technical English
- Convert text instructions into 2D graphic symbols as much as possible
- Structure the result with DITA and Information Mapping

The categorisation in phase 3 affects which actions are available. Because conceptual information is harder to visualise than procedural information, they propose to use a text optimisation strategy when converting concepts for AR screen. For their research, they have chosen ASD Simplified Technical English (STE) as a tool for text optimisation (Gattullo et al., 2019, p. 278) . For more information on ASD STE, see Section 2.1.2.

For tasks, text should be converted to visual elements, for example, annotated 2D images or drawings, videos or animations, graphic symbols that align as much as possible with familiar EN ISO safety symbols, or 3D rotatable models (Gattullo et al., 2019, p. 279).

For deciding which visual element to use, Gattullo et al. (2019, pp. 280–281) evaluate the visual elements listed above by intuitiveness for the reader, availability of the visual element type in typical industrial authoring environment, authoring effort, updating effort, standardisation, and eligibility for AR deployment. They find pros and cons with all visual element types. Drawings do not integrate or match seamlessly with real world and cause occlusion for the user. 3D models and animations take time to author, yet they suffer from easy-to-spot errors, because in AR, they are rendered directly over the real component. As opposed to YouTube or iFixit, where end-users can upload content with photos of real devices, in an industrial environment, pictures and video are hard to obtain due to access, safety, security or inaccuracies in prototype components. Additionally, when an in-development or incrementally updated component gets modified, the related multimedia must be revised.

According to Gattullo et al. (2019, pp. 281), the least of these problems occur with symbols. They are not so vulnerable to product updates, and their positioning and matching with the real component is not critical. This reduces authoring time and occlusion. Actually, the authoring time is significantly low, because the author only selects ready-to-use symbols from an existing symbol library. The drawback with symbols is that their meaning is not intuitive to recognise, requiring initial reader training before used Gattullo et al. (2019, pp. 281).

Despite this drawback, Gattullo et al. (2019, p. 281) choose symbols as the visual element to be used. However, they state that using only 2D symbols is not sufficient to, for example, locate the component to be worked on, especially without AR. Currently, annotated pictures can be used for this purpose or in the future, live camera feed with referenced graphic signs can help locating the component. A problem with symbols is that their usage lacks guidelines for technical communication. However, Scurati et al. (2018) have developed a system for using

ISO-standard symbols to represent common maintenance actions, a system that is used in the visual manual theory.

To give examples, Gattullo et al. (2019, p. 278) note that in the example in Section 2.2.2, the verbs “turn”, “pour in” and “check” will be replaced with symbols with the same meaning. The referred components involved in the task will be localised with circles and arrows. Interestingly, the result will be a diagram, to which I will return later in Section 2.4.5 in the context of multimodality. For converting the result of the task in question, text or 3D animations can be used with text optimisation. Regarding the hydraulic valve instruction in the example in Section 2.2.1, the verb “push” and the referred component (manual descent button) can be converted similarly.

Finally, reference information found from the original instruction can be organised into tables.

2.2.5 Phase 5: organise information using Information Mapping principles

Phase 5 relates to how the output should be presented to the readers after the raw content is ready. Phase 5 is an important aspect of the user experience, especially considering the multimodal possibilities of different mediums. I will return to the topic of medium in more detail when introducing my design experiment medium in Section 3.3, and when discussing the implications of my design experiment to the readers in Chapter 5, especially in Section 5.2 (user experience).

Gattullo et al. (2019, p. 281) suggest using Information Mapping to structure the layout of the visual manual for the user. They propose to keep all information, including text, images, and 3D designs constantly visible, requiring no hyperlinking where context and focus is lost in traditional paper manuals. Engelke (2013, p. 109) notes that users like to see content and context constantly available.

Gattullo et al. (2019, p. 281) propose DITA XML as the underlying data model for creating the source content, where the actual content is pure XML, thus free of form, and available for reuse and single-source publishing. When DITA XML operates as a standard for creating the source content, Information Mapping is used for designing a graphical user interface (GUI), for example, as illustrated in Figure 3:

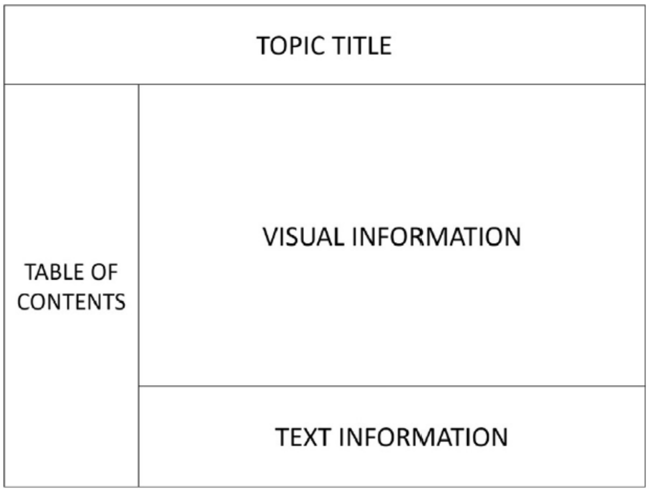


Figure 3: Visual manual GUI based on Information Mapping (Gattullo et al., 2019, p. 282)

According to Gattullo et al. (2019, p. 282), chunking described in phase 2 produces consistent and readily available content under each label of the GUI. The user remains aware of the overall procedure via table of contents (Gattullo et al., 2019, p. 282). Actually, the result bears high resemblance to the page layout (see, for example, Waller 2012), which I will discuss more with multimodality in instructions in Section 2.4.5.

2.3 Minimalism heuristics for small screen

As an established counterpart for the more modern visual manual theory described in Section 2.2, this thesis uses minimalism heuristics by van der Meij and Carroll (1998) as the second conversion theory of the design experiment. To be exact, minimalism heuristics are not a conversion theory, but rather a set of design principles for any technical communication content. Yet as discussed later in this section, they are perfectly applicable for the small screen design experiment of this thesis. There is also a practical reason: the KONE style guide that includes the guidelines for creating the current KONE technical communication content has its roots in minimalism. From the business, cost and time saving point of view, it would be beneficial not to reinvent the process and guidelines for small screen content creation if the existing process is already half way there. It would be tempting to just reinforce the current use of minimalism heuristics to encompass the small screen compatibility of the same content.

Since the conception of the minimalist instruction in the early 1980s, the concrete applications of minimalism struggled to find their way into actual instructions. To help authors apply the minimalist ideas to instructions, van der Meij and Carroll (1998) presented the design principles and their respective heuristics for designing a minimalist instruction. Based on the literary review by Virtaluoto, Suojanen and Isohella (2021, p. 8), arguments that minimalism lacks guidelines for applying it in practice on instructions still exist in the 2020s. Actually, Virtaluoto, Suojanen and Isohella (2021) propose revised minimalism heuristics as a remedy for evaluating professional technical communication instructions and processes. However, the full technical communication process is out of scope of this thesis, so for simplicity, this thesis retains to use the original heuristics in its design experiment.

According to van der Meij and Carroll (1998, p. 21), the major minimalist design principles and their heuristics are:

Principle 1: Choose an action-oriented approach

Heuristic1.1: Provide an immediate opportunity to act.

Heuristic 1.2: Encourage and support exploration and innovation.

Heuristic 1.3: Respect the integrity of the user's activity.

Principle 2: Anchor the tool in the task domain

Heuristic 2.1: Select or design instructional activities that are real tasks.

Heuristic 2.2: The components of the instruction should reflect the task structure.

Principle 3: Support error recognition and recovery

Heuristic 3.1: Prevent mistakes whenever possible.

Heuristic 3.2: Provide error information when actions are error prone or when correction is difficult.

Heuristic 3.3: Provide error information that supports detection, diagnosis, and recovery.

Heuristic 3.4: Provide on-the-spot error information.

Principle 4: Support reading to do, study and locate

Heuristic 4.1: Be brief, do not spell out everything.

Heuristic 4.2: Provide closure for chapters.

Van der Meij and Carroll (1998, pp. 21–45) explain all principles and their heuristics in detail, which I will summarise in the next paragraphs. **Principle 1** relates to the fact that instead of doing extensive learning, people learning to do new things are eager to start acting as soon as possible. **Heuristic 1.1** reverses the tutorial ideology, where you first read to learn what you will do, by making the users do first and leave the explanation later. The approach is comparable to a guided tour (van der Meij and Carroll, 1998, pp. 21–23). **Heuristic 1.2** encourages users to try out and explore undocumented features, but not totally unguided. When encouraging exploration, there is no need to document every detail. Users can apply what is instructed to perform similar tasks as long as they reach their goals. The exact route, as long as it is safe, is less significant (van der Meij and Carroll, 1998, pp. 23–25). **Heuristic 1.3** stresses that users themselves eventually select how and when they get instructed. The author or the designer of the instruction must not impose all instructions on the users, but the users' own needs must be respected and supported. Users must be given possibilities to neglect parts if they do not need them (van der Meij and Carroll, 1998, pp. 26–28).

When thinking about a software application, **principle 2** reverses the idea that all menus, features and subtleties of the application must be documented, for example, in the order they appear in the application. Instead, what gets instructed starts from the real goal of the user. Then, only the path through the application menus to the goal gets instructed. This is supported by **heuristic 2.1** by only documenting tasks that are real tasks, not something arbitrary that rarely happens when users perform real tasks (van der Meij and Carroll, 1998, pp. 28–31). Naturally, **principle 2** applies to all kinds of products, not just applications. **Heuristic 2.2** applies to the structuring of a traditional document, namely that the headings of individual tasks should support users in understanding the big picture, enabling them to see from the table of contents how the whole operation, typically consisting of individual tasks, gets completed. Properly structured headings also help the findability of a specific task when the document is used as a reference work (van der Meij and Carroll, 1998, pp. 31–33).

Even though minimalism aims at producing more compact and faster-to-get-started instructions, error prevention and recovery is something that minimalism actually *adds* to the instruction. This is described by **principle 3**. If errors happen without error recovery instructions, users spend time correcting them, which again is something that minimalism wants to avoid for helping users to accomplish their goals quickly (van der Meij and Carroll, 1998, p. 33). They argue that describing error situations and their remedies increases users' confidence and understanding of the task and the product, enabling them to detect and correct errors faster (van der Meij and Carroll, 1998, p. 35). **Heuristic 3.1** aims at preventing the errors *before* they occur with the means of clear language, short and simple sentences and by giving hints (van der Meij and Carroll, 1998, pp. 35–36). **Heuristic 3.2** continues by giving extra guidance to users during error-prone situations, for example, when the product is designed so that errors can easily occur (van der Meij and Carroll, 1998, pp. 37–38). **Heuristic 3.3** states what to do if an error occurs. The instructions must give information how to recognise and locate the mistake, understand it and how to correct it (van der Meij and Carroll, 1998, p. 38). **Heuristic 3.4** addresses the location in the instructions where the error information must be given, that is, as close to the actual step where the error potentially occurs. This is simply because the more other actions happen between the mistake and its correction, the more difficult it gets to recover from the error (van der Meij and Carroll, 1998, p. 41).

Van der Meij and Carroll (1998, pp. 42–43) state that different task performers and instruction users as humans have different reading strategies. Few read everything from the beginning to the end, some just seek a small piece of reference information, and every point in between. The instructions should try to accommodate all reading styles as best as it can. This is what **principle 4** tries to accommodate. Principle 4 may sound outdated and related to paper manuals when considering the digital, small screen focus of today's content. In modern mediums, content can theoretically be published anywhere and in any form, for example, as search results separated from their original context. However, Waller (2012, p. 238–241) notes that the *layout* often associated with traditional paper manuals is still relevant as a *utility for aiding readers*, especially for strategic readers. Strategic readers are those studying and learning, or reading user manuals, and whose requirements for the content are not simple and linear. According to Waller (2012, p. 238), many current digital devices are a better fit for linear texts, for example, fiction. For strategic readers, bullet lists, diagrams, step lists, or any layout formatting provide structural cues that operate as memory tools when moving between parts of content non-linearly (Waller 2012, p. 241) . I will return to the *page layout* when discussing the theories of multimodal communication within the context of instructions in Section 2.4.5.

Heuristic 4.1 encourages to be brief in both sentences and chapters, for example, not to give explanations for the sake of explanations. To be brief signals the users that the task that they are attempting is not a difficult one, does not require overbearing effort and helps them to perform by encouraging their own thinking (van der Meij and Carroll, 1998, pp. 43–44). **Heuristic 4.2** instructs to provide clear starts and closures for each chapter. For readers using the manual as a reference work in random order, or if starting from the middle, the independently starting and closing chapters ease the findability, task performance and completion of tasks (van der Meij and Carroll, 1998, pp. 44–45). The thinking is the same as with modular design and DITA XML. For more information on modular design and DITA XML, see Section 2.1.1.

2.4 Theories of multimodal communication

So far in Chapter 2, I have discussed technical communication and its standards and theories in the context of the design experiment of this thesis. Two of them, visual manual and minimalism heuristics, were given a more detailed focus, since I will use them to convert existing paper format instructions to the small screen of smart glasses in Chapters 3 and 4. But the question is, if authors simply follow such theories, will the resulting small screen content automatically be good and understandable? This is a vast question and this thesis can only scratch the surface, with many aspects falling out of scope of this thesis. But no matter how the content gets authored, would it be good to at least have tools for analysing how the readers understand the instructions, and especially why? Bateman (2008, p. 6) argues that even when a document has been authored well, it is not guaranteed that the readers will recognise and interpret the interplay of all visual aspects working together on the page of a document. This is where *theories of multimodal communication* can help, and the reason why I am using them to analyse the results of the conversions.

Bateman (2008, p.7) argues that traditional linguistic views of text nor traditional views of the image do not provide the meaning of *interplay* described above within them. When reading a document of any kind, Bateman (2008, p. 5) notes that readers must work out what is, but also, what is not a sensible interpretation of the words, figures and whatever is presented in various ways on the page of a document. Today, the intended meanings of such a *multimodal* document are distributed over a broader range of visual *modes* than some 30 or 40 years ago (Bateman, 2008, p. 6). I will return to *modes* in detail in Section 2.4.2, but let us first briefly set the focus back to the problem setting of this thesis.

To orientate why theories of multimodal communication matter to Industry 4.0 small screen instructions, imagine a maintenance technician in an elevator shaft, or in any small technical space, surrounded by potentially dangerous mechanical and electrical equipment. With today's rapidly evolving technology, the technician must perform a repair task without previous experience of the equipment, meaning that instructions are required. What would the best way to transmit the instructions be, and more importantly, what would the instructive content in the transmitting medium be like to optimally communicate the instructions for the technician? For the technician to complete the task successfully, just to have some instructions is not enough. The instructions should be easy to understand, maximally accessible, and their implementation should minimise all risks of injury to the technician. There are many questions and requirements, so a systematic approach is required. Technical communication needs theories and methods for analysing communication more now than ever, especially when new media (AR) are being taken into use.

To broaden the view, let us consider human–computer interaction, where multimodality is primarily understood in terms of senses, such as sight, hearing and touch, Siltanen and Heinonen (2020, p. 107) call for more multimodal digital user interface development and further research on the area as an answer to varying use situations and users' personal preferences. They state that the tasks that the technicians perform vary from hands-free to hands-busy situations, when they are holding tools and equipment. Also, often the space they are in limits the use or some sensory channels of the user interface, for example, noisy environment can disturb voice commands and aural interaction, and small spaces can limit operating user interfaces with hand gestures. Siltanen and Heinonen (2020, p. 107) conclude that selectable sensory modalities of the used medium are important to match the space and person. Also, equally important is the semiotic, multimodal aspect of the communicative context in which the situation described above takes place: the sensory modalities must be paired with appropriate semiotic modes to present the content effectively. We need to cover the entire situation at hand.

To elaborate the scenario above, not only the multimodal affordances of the used small screen device are important, but the semiotic, multimodal qualities of the content that gets displayed on the small screen are crucial. That is, no matter whatever the instructions are presented on and how they may be interacted with (user interface), the content is equally important for the outcome of the task. As much as the technician benefits from the latest innovations of modern small screen devices, for example, content interactivity, visual and aural feedback, haptic feedback, voice and gesture commands and so on, the content and how it is expressed still determines whether the instructions are understandable or not.

Given the goals set out for this thesis, an in-depth study of sensory modalities and their role falls outside of the scope of this thesis. Furthermore, it must be noted not to mix the human–computer interaction definitions of multimodality with the semiotically-oriented definitions, as they are not entirely compatible (Bateman, Wildfeuer and Hiippala, 2017, pp. 41–43). Thus, it is important to underline that this thesis focuses on the semiotic aspects of multimodal communication, even when analysing the use of smart glasses. However, since RealWear HMT-1 smart glasses, a medium referred to by Siltanen and Heinonen above, are used in the design experiment of this thesis, the human–computer interaction issues are touched upon if they influence the content, its presentation or understandability. After all, Bateman, Wildfeuer and Hiippala (2017, p. 114) note that both the material side of the medium and the semiotic modes deployed on the medium must be account for when analysing a communicative situation.

2.4.1 Historical perspectives on multimodal communication theories

Contemporary theories of multimodal communication are rooted in the fields of linguistics and semiotics, as both are concerned with making and exchanging meanings. Bateman, Wildfeuer and Hiippala (2017, pp. 51–63) summarise the influence of the two founding fathers of semiotics, Ferdinand de Saussure and

Charles Sanders Peirce on multimodality: the three-part sign interpretation of signs by Peirce is closer to multimodality than Saussure’s more linguistically motivated two-part sign interpretation, that is somewhat limited by its linguistic focus. Peirce, on the other hand, was never limited to language, and many of his semiotic innovations went beyond verbal signifying practices (Bateman, 2018b). As opposed to the two-part sign interpretation of Saussure, where a sign consists of an arbitrary signifier and the object of what is being signified, the three-part interpretation of Peirce puts less weight on the arbitrariness of the sign in the linguistic fashion by de Saussure. The three-part interpretation accepts all kinds of signifying material, thus being more useful with pictures, film, and other media that interests us multimodally. In the three-part sign interpretation, we have the object, interpretant and the representamen. Figure 4 illustrates the models of the sign by Saussure and Peirce:

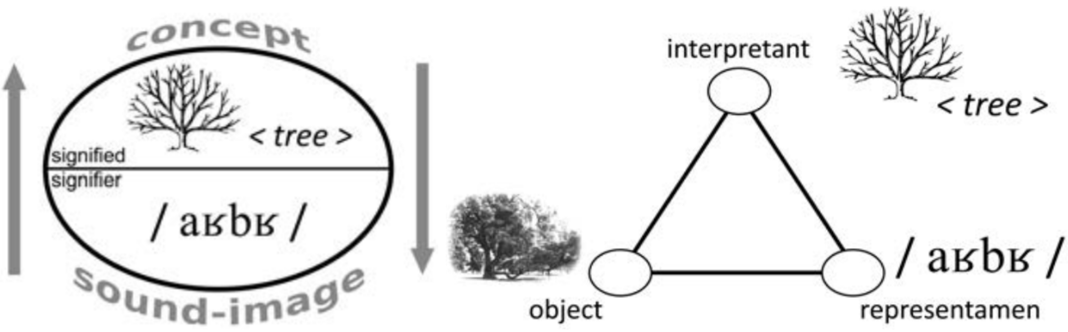


Figure 4 Saussure’s two-part model of the sign (left) and Peirce’s three-part model of the sign (right) (Bateman, Wildfeuer and Hiippala, 2017, pp. 54–57)

In Peirce’s model, “<tree>” is only a suggestion that trees are at issue. However, this should not be understood in only linguistic sense, because it could be anything, for example, an image of trees swaying in the wind or the smell of the leaves (Bateman, Wildfeuer and Hiippala, 2017, p. 57). See, for example, Jappy (2013) for Peircean semiotics in the context of multimodal documents in our modern visual culture.

Providing a comprehensive review of literature on multimodality is not easy, because the field is extremely broad and not yet fully developed. Bateman, Wildfeuer and Hiippala (2017, p. 71) note that the central idea of multimodality itself, combining different forms of expression, is old. We can find examples of multimodal communication ever since humans started manipulating materialities for communicative purposes. Bateman, Wildfeuer and Hiippala (2017, p. 23) and Wildfeuer et al. (2019) doubt that there is an established academic field of multimodality. The concept of multimodality has been widely used in semiotics, discourse and media analysis, but at the risk of being worn out and thus too a fuzzy label (Deppermann, 2013, p. 2). Bezemer and Kress (2016, p. 142) state that multimodal studies appear in various disciplines, with different theoretical and methodological frameworks in each. Like Bateman, Wildfeuer and Hiippala (2017), Bezemer and Kress (2016) aim to provide an approach to unify different domains and materials, because what humans are actually trying to achieve in general is to make meanings, regardless of whether the label is communication, learning, or a disciplinary approach, such as sociology, psychology, anthropology, pedagogy, linguistics or semiotics. All signs need to be considered equally, no matter who are the sign makers or what is the mode of communication. What must be examined and understood is the role of the interest of sign-makers, and how the path unfolds through the principles at work in transformation, interpretation and ultimately in meaning-making to sign consumers (Bezemer and Kress, 2016, pp. 5–6).

To analyse forms of multimodal communication, various tool kit approaches exists with ready-made distinctions to apply to images, text, sound and so on, but when confronted with real-life communicative situations, which are inherently multimodal, the situation can be so complex that the tool kit model cannot be fully applied (Bateman, Wildfeuer and Hiippala, 2017, p. 1). To tackle and analyse the real-life multimodality, Bateman, Wildfeuer and Hiippala (2017, p. 1) offer multimodal competence to build your own tool kit for the exact multimodal situation at hand. After all, it is better to analyse what you are analysing instead of just describing what you are analysing (Bateman, Wildfeuer and Hiippala, 2017, p. 231). For this reason, I will analyse the results of my design experiment according to Bateman, Wildfeuer and Hiippala (2017) in Section 4.3.

2.4.2 Semiotic modes as the essence of multimodality

Semiotics and semiotic modes are the foundation of multimodality (Bateman, Wildfeuer and Hiippala, 2017, p. 112). To understand multimodal communication, one must understand what is meant by a semiotic mode. Let us consider an illustrated page of a book with text, images, diagrams or charts. It can as well be a modern, digital information product of any kind. Bateman (2008, p.1) states that nowadays there are more components than just text, because the visual aspects (illustrations) are interwoven around the text, or sometimes even replacing the text. These visual aspects, also including the text, are the modes of information presentation. Bateman (2008, p.1) states that:

> "When combined on the page, we have a multimodal document. In such artefacts, a variety of visually-based modes are deployed simultaneously in order to fulfill an orchestrated collection of interwoven communicative goals."

Therefore, when multiple semiotic modes are operating together, we have multimodality. To broaden the view for understanding how widely multimodality is present, Bateman (2008, pp. 7–8) notes that a combination of modes is typical in the presentation of today's advertising, education, and entertainment products, among others. Even though these are delivered using different technologies, each of which can affect which modes can be shown on the technology in question, the semiotic modes are present on content regardless of the delivery technology.

Semiotic modes are more detailed and granular than what are traditionally thought of as modes, such as the visual mode, sound or language as a mode. Furthermore, not all semiotic modes are even known yet, but more will be found through empirical analysis (Bateman, Wildfeuer and Hiippala, 2017, p. 136). To understand semiotic modes, one must realise that they are more than just the sensory channels. Bateman, Wildfeuer and Hiippala (2017, pp. 27–28) state that:

> "Sound is not just sound or the hearing of tones; it also gives information about space, hardness, distance and direction… When we turn to the use of sound in language, the situation is even more complex. We have all of the physical properties we have described so far plus indications of emotional state, physical state (age, health, tiredness) and much more—even without considering the particular phonetic details of speech."

According to Bateman, Wildfeuer and Hiippala (2017, pp. 113–114), literature agrees that modes have a material dimension that relates back to the sensory channels perceiving them. They continue to state that also the material itself has a semiotic dimension, since the material always has some kind of significance to its users. Yet, opinions disagree how the material and the semiotic are related. According to Bateman, Wildfeuer and Hiippala (2017, pp. 113–114), this can be tackled by recognising that semiotic modes have both the material and the semiotic dimension (see also Bateman 2011). The material side determines the canvas (see Section 2.4.3) and its affordances where the signs are made on. The semiotic side determines which signs in that material make meanings and in which mode. Together, these constitute the meaning of a semiotic mode (Bateman, Wildfeuer and Hiippala, 2017, pp. 113–114).

To sum up the case for Industry 4.0 instructions and the design experiment of this thesis, it is important to analyse the spatial surroundings in which the communicative situation takes place (elevator shaft), the medium used (screen of RealWear HMT-1 smart glasses) and the redesigned instructive content (visual manual and minimalistic versions) as constituting one multimodal user experience, where the goal is to understandably signify the technician to perform the right things and get the task completed. In the next sections, I will consider aspects of multimodal theories that are beneficial for analysing Industry 4.0 instructions.

2.4.3 Genre, medium and canvases

We readily identify different types of films, books, magazines and whatnot by their appearance and content. Bateman (2008, p.9) states that genre is important for multimodal analysis, because it provides the point of *comparison*. The meanings we make in any communicative form or genre are based on the history and societal context of the object (for example, document). Thus, Bateman (2008, p.10) argues that interacting with different objects is affected by the resources that have been used with them earlier. In other words, objects of different genre are authored and interpreted differently. For example, the mark “x” on a treasure map carries a different meaning than the “x” in a mathematical equation.

I will return to historical and cultural considerations in Section 2.4.4, but let us now focus on the role of materials in communication, for example, film, book, or the treasure map (the medium). According to Bezemer and Kress (2016, p. 31), sign makers make their signs using the material affordances of modes. They state that the material’s resources shape the arrangement of modes, for example, images support spatially organising marks on paper, speech utilises the physics and rhythm of breathing, and gestures require spatial and temporal nature of action.

Bateman et al. (2017, p. 109) attempt to impose control over these phenomena by defining media that further provide *canvases* on which semiotic modes may be deployed. To understand the affordances of mediums and their canvases in relation to participating semiotic modes, Bateman, Wildfeuer and Hiippala (2017, p. 124) state that meaning cannot be realised in a medium. Meaning can only realise in the *semiotic mode* participating in the medium. Bateman (2008, p.16) states that, for example, documents are actually produced on a canvas (not on a medium). The canvas can be a specific paper type, or a specific digital screen type. The bottom line is that each canvas has its own specific properties that affect what can be placed on them. Regarding the likes of RealWear HMT-1 smart glasses used in this thesis, Bateman, Wildfeuer and Hiippala (2017, p. 126) note that a medium is more of a communicative form than physical material. Therefore, for example, the screen of a tablet computer is not a medium in that sense. Instead, it is a virtual canvas employed by a variety of media. In the context of smart glasses, the screen of the smart glasses is a canvas comparable to a tablet computer and the canvas is a way of breaking down the medium into smaller and manageable pieces. The canvases provided by the smart glasses extend beyond the immediate view of a single step in an instruction, as alternative views (the following steps, see Section 4.2 for examples). The alternative views (canvases) can be rendered by manipulating the medium via the user interface.

To keep up with the modern media, one must be ready to analyse the multimodal situation as it is and not to force it to any pre-defined category. Therefore, rather than attempting to fit Industry 4.0 instructions into any pre-defined categories, this thesis builds the analytical apparatus from the ground up, using the framework proposed by Bateman, Wildfeuer and Hiippala (2017). To understand the medium utilised in this thesis, Figure 5 by Bateman, Wildfeuer and Hiippala (2017, p. 109) illustrates a classification of the mediums that can be used in communicative situations:

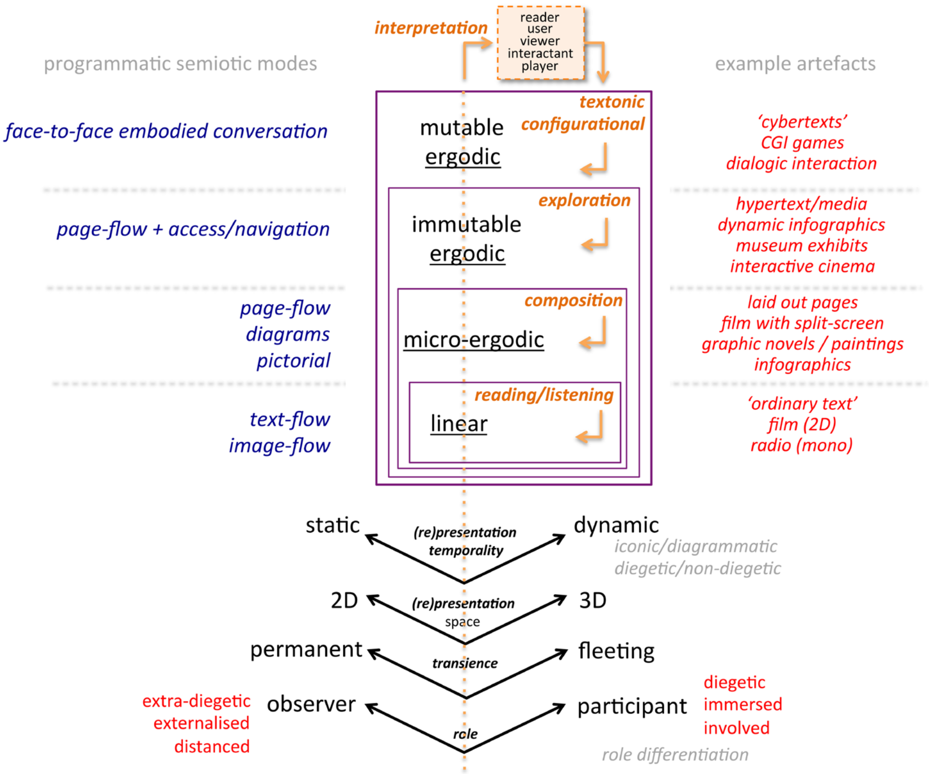


Figure 5: Classification of medium in communicative situations (Bateman, Wildfeuer and Hiippala, 2017, p. 109)

Bateman, Wildfeuer and Hiippala (2017, pp. 109–110) describe Figure 5 in detail, which I will summarise in the next paragraphs. Starting from the bottom, the lowermost distinctions (role, transience, presentation space, presentation temporality) are conceptualised as the material properties of the medium. For example, in a face-to-face discussion, the sign consumer is a participant in a fleeting conversation that happens in 3D space with, for example, gestures and dynamics of sound as the modes of meaning-making. Traditional instructions take the other route here, where the sign consumer is an observer, reading permanently printed text and pictures on a static 2D paper (Bateman, Wildfeuer and Hiippala, 2017, pp. 109–110).

The central boxes describe the amount of work that the reader must perform to extract the meaning out of the object of interpretation. Linear refers to communicative situations taking place on canvases that unfold in one dimension, for example, reading traditional, linear-interrupted printed text, or listening to a radio show or a podcast. On micro-ergodic canvases, the sign-consumer has a choice from where to start approaching what is presented similarly to, for example, starting to explore a map. Another example is a diagram, which is a typical component in technical communication content. For example, a diagram in IKEA instructions makes use of the two-dimensional spatial extent provided by the canvas. On immutable ergodic canvases, for example, a web page navigation, the sign consumer can freely choose the order of approaching what is presented, but cannot modify the content. This is the equivalent to interacting with the RealWear HMT-1 user interface. On mutable ergodic canvases, the sign consumer can actively influence the sign making process itself. For example, face-to-face conversation is a classic example, in which the participants dynamically negotiate how the communication proceeds. The boxes are positioned inside each other to signify that communicative situations taking place on canvases may be embedded within one another (Bateman, Wildfeuer and Hiippala, 2017, pp. 109–110).

Example media of the boxed media types are listed on the right-hand side, and potential semiotic modes appear on the left-hand side. The upward dashed line and arrow is the sign consumers’ path of interpretation, and downward arrows are strategies that the sign consumers need for making meaning of what they experience (Bateman, Wildfeuer and Hiippala, 2017, pp. 109–110).

To map Figure 5 to the design experiment of this thesis, Figure 6 starts by illustrating a simple definition of the terms *medium* and *canvas* in the context of this thesis:

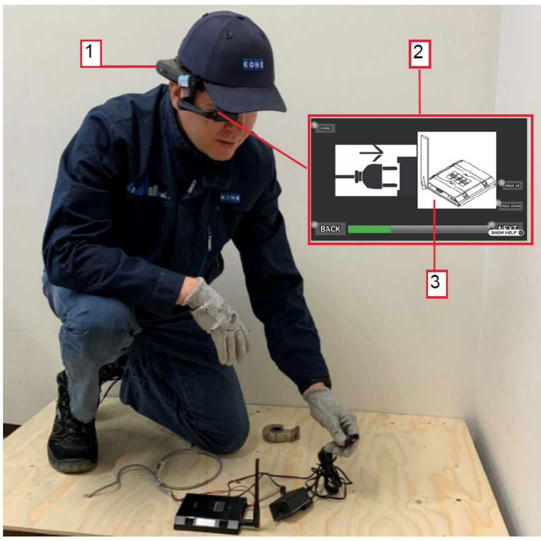


Figure 6 Medium and canvas in the context of this thesis

Here, the concepts of medium and canvas can be seen from different angles, thus affecting the analysis. From the point of view illustrated in Figure 6, the screen (2) of

RealWear HMT-1 smart glasses (1) provides the underlying materiality or technical medium. The medium makes available an extensible, dynamic canvas (3), which hosts different semiotic modes. These semiotic modes carry the actual content responsible for technical communication. Figure 7 illustrates a wider approach, capturing the multimodal *situation* in the scope of this thesis:

Figure 7: Multimodal situation: technician working on simulated elevator car roof according to AR instructions

In the situation above, there are many multimodal elements in operation. The spatiotemporal dimensions of interacting with external media (instructions on smart glasses) on top of an elevator car roof (1) is the overall embedding communicative situation, in which everything else takes place. The voice-controlled, immutable ergodic navigational user interface of the smart glasses and its applications (2) is one element, the table of contents -type contextual information that the user would use for finding, selecting and navigating between and in the instructions (3) is another, and finally, the actual instruction screens (steps) displayed on the screen (4) provide its own element. So, the medium in which the design experiment of this thesis is embedded is interaction with an external medium (smart glasses) in an extremely constrained physical environment, that is, the elevator shaft. From this angle, there are many possible canvases where signs can be made on for the technician to interpret: the physical space of the elevator shaft, the smart glasses, the space and layout of the actual instruction content, and even the maintained equipment with its possible warning stickers can be seen as one.

2.4.4 Social semiotics and the relationship between culture and multimodality

An interesting aspect of meaning-making in global technical communication, from the point of view of the sign maker (author of instruction), sign (content of instruction) and sign consumer (reader of instruction), is provided by social semiotic theories of multimodality. They emphasise the cultural dimension of multimodality. In other words, how social and cultural differences affect in the multimodal meaning-making process. Social semiotics take signs as elements in which the signified, the meaning, and the signifier, a material form, appear together (Bezemer and Kress, 2016, p. 9). The multimodal semiotic world is a social world, where power and location of people shape our actions and our processes of meaning-making. Here, material and non-material are unevenly distributed, also from the social point of view. Factors that matter are, for example, work and profession, gender, generation, cultural and linguistic background, education, and even class. Also notable is the factor of semiotic affordances, that is, a sign-maker's work is shaped by what others have done before in response to similar needs (Bezemer and Kress, 2016, p. 31). As a result, semiotic resources are socially shaped and specific to different cultures (Bezemer and Kress, 2016, p. 25).

Despite the importance of social and cultural differences affecting the multimodal meaning-making, these are not taken into consideration in the design experiment of this thesis to limit the scope of this thesis. The analysis of this thesis will be made from the western European social and cultural point of view. Yet, it would be interesting to study further if and how Industry 4.0 and digital globalisation effect the social and cultural differences in the context of technical communication. As Bezemer and Kress (2016, p. 31) state, semiotic affordances are constantly re-shaped.

2.4.5 Multimodality in instructions

Bezemer and Kress (2016, pp. 5–6) state that it is not important what discipline is analysed, since we aim to understand the meaning-making process. With this thought, we can safely assume that instructions need not to differ from any other branch of multimodal analysis, so we do not have to look for specific methods for analysing instructions multimodally. However, some distinctive features can be found from instructions, which this section will address. As Bezemer and Kress (2016, p. 81) emphasise, we need to understand the unique factors in each environment, with the expanding digital or virtual ways of learning.

Bezemer and Kress (2016, pp. 33–34) discuss the availability of modes, which simply means that, for example, using a video or an animation is not available on a printed A4 instruction paper. These count as constraints on sign-making. Regarding printed IKEA instructions, Bezemer and Kress (2016, p. 81) note that, for example, gesture is not available to signify what should be done. The sign consumer is limited to image and writing, which provides different potential in multimodal meaning-making. In the design experiment of this thesis, this constraint clearly applies to the A4 paper-based original instructions, but would not necessarily need to restrict the redesigned versions, especially considering that the medium, the smart glasses, have a digital small screen that is comparable to a phone or a tablet operating system. The choice of one mode rather than another affects what can be and gets communicated (Bezemer and Kress, 2016, p. 36). Therefore, exploring all modes available through the smart glasses might lead to using different modes in the meaning-making process. This subject gets discussed more in Section 5.3 that identifies areas for further research.

To understand what to expect when analysing instructions in the context of this thesis, let us take a top-down consideration of a traditional instruction. A traditional instruction is often assimilated to have a *page layout*. Waller (2012, p. 242) notes the significance of layout for different genres, for example, newspapers, school textbooks, and technical communication. According to Waller (2012, p. 242), the layout immediately signifies the genre to the reader, for example, engaging layout and headings invite newspaper or magazine reader to browse, and the ordered and controlled layout of an instruction, with diagrams, large headings and step list numerals, messages systematic reading.

Theories of multimodality can be used to granularly analyse how the *page layout* is constructed and why it is so effective as a resource to carry meanings. Figure 8 by Bateman (2008, p. 106) identifies the semiotic modes working on a page:

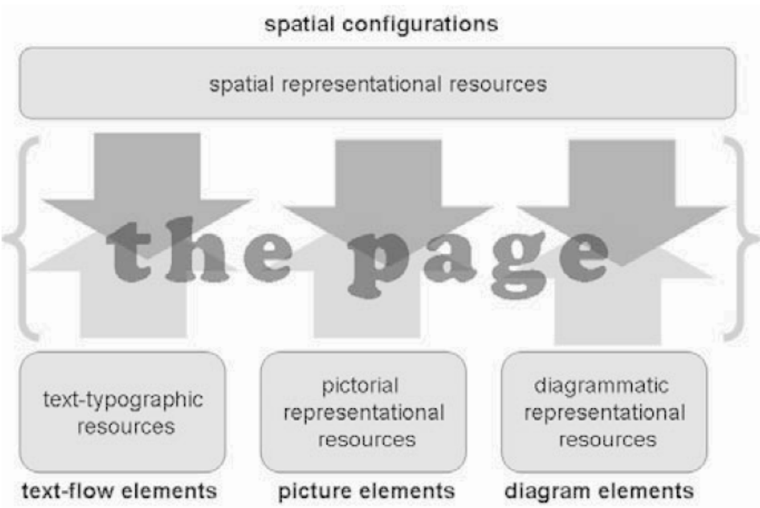


Figure 8: The page as a site of cooperation and integration of distinct semiotic modes (Bateman, 2008, p. 106)

To break down Figure 8, Hiippala (2015, p. 16) explains the spatial representational resources, the layout, which can take and combine contributions from other representational resources (text-typographic resources, pictorial representational resources and diagrammatic representational resources) and combine them on the page. The layout itself can also be considered as a resource, because it combines the other contributions to equal more than the sum of their parts. To work forward in Figure 8, with the text-typographic resources and pictorial representational resources, Matthiessen (2007, pp. 24–25) summarises the semiotic systems of the printed page to consist of the written language that can be read aloud, visual paralanguage consisting of font family, type face ("style"), layout (graphic design) and visual (pictorial) semiotic systems, for example, drawings, paintings, photographs, maps, graphs, charts, and so on.

To progress to the semiotic modes in operation on a page, Figure 9 by Bateman (2008, p. 175) illustrates the three semiotic modes that are commonly found from the pages of multimodal documents: page-flow, text-flow and image-flow.

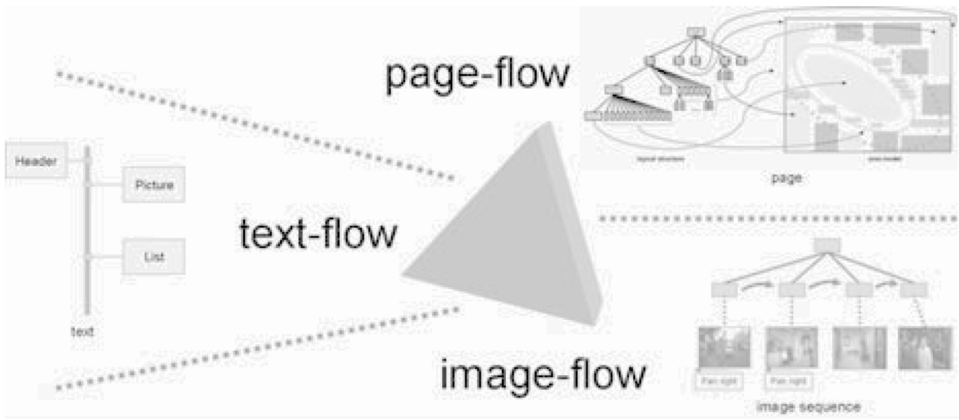


Figure 9 Semiotic modes common on document page (Bateman, 2008, p. 175)

Bateman (2008, p. 175) states that the semiotic mode of **text-flow** is linear, one-dimensional, written text, which can also include figures, pictures, diagrams, or other modes, close to the text that refers to them. Here, it is important to understand that text-flow is not only text (see Section 4.2.1 for my text-flow analysis of the design experiment of this thesis). As seen in Figure 9, **image-flow** constitutes of linear graphic sequences, one after another, as opposed to text (Bateman, 2008, p. 175). Comics are a good example of image-flow. According to Bateman (2008, p.176), **page-flow** utilises the page's whole spatial area, two-dimensionally, for making meanings. Again, it is important to notice that page-flow can combine elements from other semiotic modes to the same page, for example, text-flow, diagrams and figures (Bateman, 2008, p.176).

To continue the top-down journey of the page in instructions, one cannot avoid the diagrammatic representational resources. A well-known example of technical communication manuals are IKEA instructions, which Bezemer and Kress (2016, pp. 82–83) and Bateman, Wildfeuer and Hiippala (2017, pp. 286–288) analyse as mainly using the diagrammatic mode (assembly drawing with additional lines, zoom bubbles, screws and screwdriver markings superimposed on the actual drawing of the furniture), and the semiotic mode of comics, for example, the cartoon characters calling IKEA in case of problems.

The diagrammatic mode seems to be widely used in instructions, judging from examples in multimodal literature and from the material in this thesis. Bateman, Wildfeuer and Hiippala (2017, p. 279) name typical examples of the diagrammatic mode as charts, graphs and schematic drawings, which can combine with other semiotic modes in many other media. For example, diagrammatic elements can be combined with photographs, illustrations and drawings (Bateman, Wildfeuer and Hiippala, 2017, p. 294). As examples of page-based media, they mention scientific articles, school textbooks and assembly instructions. They also note that the diagrammatic mode combines easily with other semiotic modes, forming composite units. Figure 10 that is taken from the design experiment of this thesis illustrates an example of the diagrammatic mode:

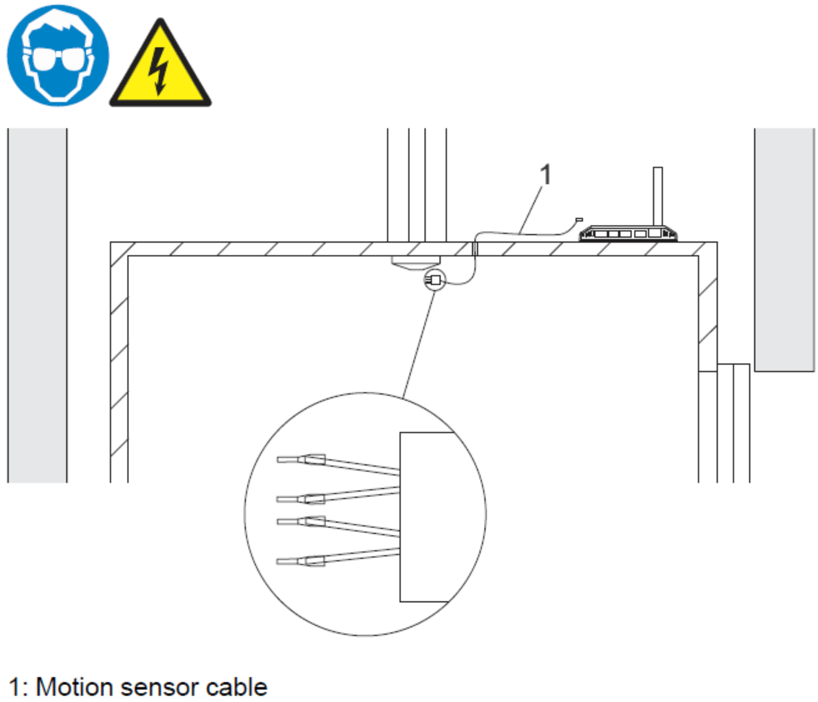


Figure 10:Drawing of elevator car, connectivity device and wiring in elevator shaft with diagrammatic elements (KONE instructions)

To the drawing, the diagrammatic mode adds symbols, a callout number (1) with corresponding text at the bottom and connecting lines with zoom bubbles that point to specific parts of the drawing. In assembly instructions, the diagrammatic mode adds many visual semiotic resources (Bateman, Wildfeuer and Hiippala, 2017, p. 280).

Figure 11 illustrates the diagrammatic mode theoretically (Hiippala and Bateman, 2020, p. 3):

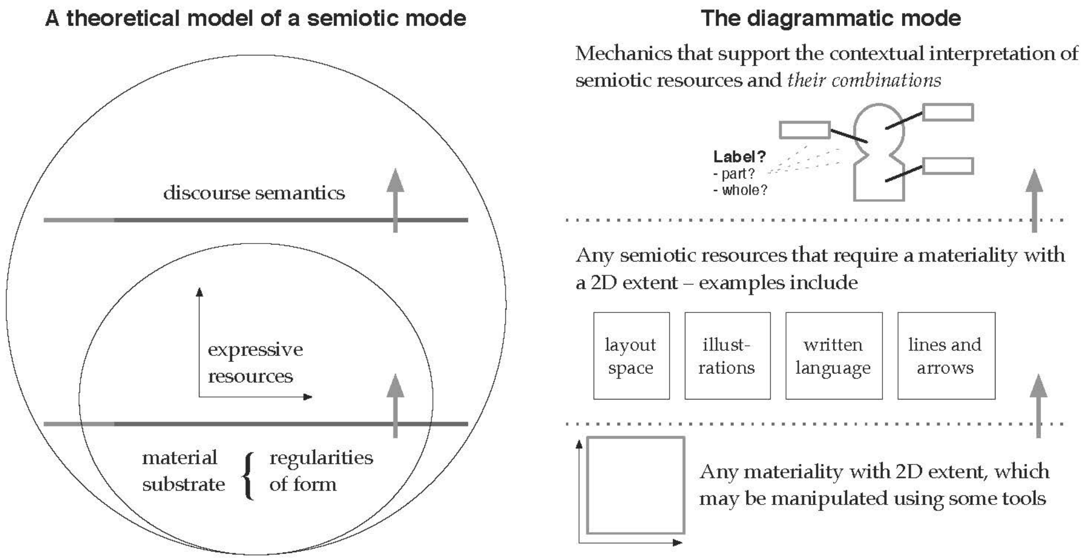


Figure 11 A theoretical model of a semiotic mode and a sketch of the fundamentals for a diagrammatic mode (Hiippala and Bateman, 2020, p. 3)

Starting from the bottom, the diagrammatic mode is a semiotic mode, which requires a materiality (medium), where the expressive resources can be placed on. In this thesis, the medium is the small screen of RealWear HMT-1 smart glasses. In the middle of Figure 11, the diagrammatic mode uses expressive resources for making signs for the readers, such as text, pictures, zoom bubbles, callout numbers, and the layout of the medium, for example, the page layout. Thirdly, discourse semantics are used to explain the communication and contextual interpretation that the reader of the diagram makes. It is the discourse semantics that explain why the expressive resources may receive different interpretations between different contexts, including different readers (Hiippala and Bateman, 2020, p. 3).

Another typical genre in technical communication is the information graphics mode (Bateman, Wildfeuer and Hiippala, 2017, p. 289). Typical information graphics can be found from newspapers and magazines, where, for example, a relatively large space of a newspaper opening is used to present topics such as the results of a parliament election. Here, the support for different political parties can be presented with a map of election areas, colours, text, embedded tables, and nested graphics and charts. When reading diagrams or information graphics, the reader has to work through the layout and nested graphics to understand how they relate to each other (Bateman, Wildfeuer and Hiippala, 2017, pp. 283–284). The difference between a diagram and an information graphic is that a micro-ergodic information graphic can be freely explored to arouse the readers' curiosity, whereas a diagram simply seeks to instruct how to assemble a product, and aims to downplay the ergodic dimension for maximising the readers' success to reach their goals (Bateman, Wildfeuer and Hiippala, 2017, p. 286).

Finally, relevant for Industry 4.0 instructions are YouTube-type video instructions and animations (Bateman, 2018a). They provide additional semiotic modes as compared to traditional paper-based instructions. Moving from still image-based traditional learning environments (instructions) to a tutorial in video and speech is a significant shift in semiotic work. Suddenly, we have new semiotic resources and modes in effect to be analysed, ranging from acting and speaking, considering shooting angles, focus, and framing, adding written captions and music, and cutting (Bezemer and Kress, 2016, p. 84).

However, whether videos and animations should or should not be used more in instructions, or whether they perform better or worse in the technical communication meaning-making process, is not in the direct scope of this thesis, simply because the material in the design experiment did not include them. I also chose not to add them to the conversions, because the researched theories of visual manual and minimalism heuristics did not require using them. Despite of this, their potential was recognised by the test users, and I will return to this topic in the discussion in Chapter 5. The more ergodic navigational type of multimodality is also out of scope of this thesis, for example, online help and web page type of navigation. The multimodal focus of this thesis is on the content communicated on a single page. However, this thesis notes the ergodic nature of the smart glasses as a physical medium.

3 Methods and materials

This chapter presents the methods and materials used in this thesis, including descriptions of the technical devices utilised in the user test I conducted with my KONE colleagues. Section 3.1 specifies the selection criteria for the material used in the design experiment, and focuses on the methods derived from the theoretical framework of this thesis (see Chapter 2). Section 3.2 describes the design experiment and user test in detail by introducing the used materials (instructions) and describing the user test setting for which I created the small screen conversions of the instructions. The small screen of the design experiment, RealWear HMT-1 smart glasses, is introduced in detail in Section 3.3, while Section 3.4 wraps up the design experiment and user test description by introducing the physical surroundings, the multimodal situation, where the users of the instructions work in real life and in our user test environment.

3.1 Multimodal methods

Systematic sampling was used to select the data from the KONE document library for analysis. The criteria were to select instructions that benefit from augmented reality, have potential for Industry 4.0 use (include new digital equipment), and could be tested in a safe environment with real users with different backgrounds and experience. The selected materials are introduced in section 3.2 and the full array of screenshots is visible in the analysis in Section 4.2.

This thesis combines experimental and qualitative methods by having authored versions of task instructions (the data), whose design is theoretically-informed (enriched with visual manual and minimalism heuristics). For this data, a qualitative analysis is conducted of its simulated use through a design experiment. Any quantitative results from the user tests conducted with my KONE colleagues can be analysed in future research. To support the qualitative analysis, the data is transcripted as screenshots of each step of the instructions. All the data is transcripted for the design experiment part, but due to the similarity of the data, a full multimodal analysis is conducted only for one representative step. The analysed step is step 5, "move SIM card", of the "Replace KONE MediaPlayer" task. For more information of the selected step, see Sections 4.2.3 and 4.3. However, Chapter 5 discusses the whole data also from the multimodal perspective, to see how redesigning the data according to the principles of visual manual and minimalism heuristics affect the data multimodally.

3.2 Evaluation through design and user test experiment

To get data for the multimodal analysis, this thesis conducted a design experiment as an evaluation method to convert three originally A4 paper format short task instructions according to the visual manual and minimalism heuristics theories. The conversions were published to RealWear HMT-1 smart glasses for analysis and discussion, and for qualitative discussion with test users.

The user test with 21 test users was conducted in cooperation with my colleagues at KONE Oyj. Test users were observed performing the tasks (the data of this thesis) on a test setup. They answered a questionnaire to evaluate the task versions. The quantitative results (the questionnaire) of the user evaluation test are out of scope of this thesis, but the preparations, such as creating the test environment to mimic real equipment in real environment, testing the instructions myself in the test environment, and qualitative comments from test users are included in the analysis and discussion sections. Including the user evaluation test preparation and qualitative test user feedback adds validity to the results of the design experiment, because the created task versions are thus proven to be designed, created and tested to be used in real environment.

The target of the design experiment was to find out how the selected theories would perform in practice, to see the real outcome on a real small screen, and to evaluate the whole process from the point of view of a technical communication professional. Secondly, the target was to see the multimodal opportunities that are and could be available with the same materials, and how theories of multimodality could be used to analyse the results of the conversions.

The instructions used as the data in the design experiment are authentic installation and maintenance tasks that are in active use at KONE Oyj. They contain instructions for replacing a content streaming device (KONE MediaPlayer) and installing a connectivity device (KONE Connection) into elevator environment.

The following instructions form the data of my design experiment:

- "Install motion sensor cable"
  The user goes on the elevator car roof and routes a motion sensor cable through the elevator car roof into the elevator car.
- "Prepare KONE Connection"
  On the elevator car roof, the user connects the motion sensor cable of the previous task with a splitter wire to KONE Connection, switches on KONE Connection and checks that KONE Connection boots up correctly.
- "Replace KONE MediaPlayer"
  The user replaces a broken KONE MediaPlayer behind a lobby TV.

For full versions of the original PDF instructions, see Sections 4.2.1, 4.2.2 and 4.2.3. Figure 12 illustrates each of the tasks in the test set up we created:

Figure 12: Technician performing the tasks "Install motion sensor cable" (left), "Power up KONE Connection" (middle) and "Replace KONE MediaPlayer" (right)

Due to the small screen limitations of the RealWear HMT-1 medium, all task versions required at least some modifications for showing them on the screen, even the original. Because of the modifications, the "original" version is called "conventional". The detailed modifications are discussed in Chapter 4.

Out of the three instructions described above, I produced the following three conversions:

- Conventional
  This version is included for comparison to illustrate what the content looks like without any theoretically-informed redesign.

- Visual manual
  This version was produced according to the visual manual theory by Gattullo et al. (2019). For more information on the visual manual theory, see Section 2.2.
- Minimalistic
  This version was produced according to minimalism heuristics by van der Meij and Carroll (1998). For more information on minimalism heuristics, see Section 2.3.

The results of the converted versions are visible in Sections 4.2.1, 4.2.2 and 4.2.3, including the original A4 PDF version. As a result, nine instructions were published on RealWear HMT-1 smart glasses.

All conversions were authored in DITA XML, a widely-used standard in technical communication. DITA XML is also the suggested format in the visual manual theory (Gattullo et al., 2019, p. 278).

In the following sections, I will describe the RealWear HMT-1 medium and the physical user test environment in more detail. For more information about *medium* and *canvases*, see Section 2.4.3.

3.3 RealWear HMT-1 medium

Before discussing the various tasks and their conversions, it is important to understand the properties and affordances of the medium on which the conversions are made for: the head-mounted RealWear HMT-1 smart glasses. The glasses run the Android operating system, and are designed for industrial workers with hands-free voice-control, allowing the users to perform their tasks hands-free (RealWear, 2019). The display of RealWear HMT-1 truly qualifies as a small screen: the size of the screen is only 8.4 mm, with a resolution of 854 x 480 pixels. For further multimodal applications, it has a video and still camera, loudspeaker, headphone port, four digital microphones with active noise cancellation for voice recognition, and a LED flashlight. For connectivity and sensors, it has Bluetooth, Wi-Fi, GPS, accelerometer, magnetometer and gyroscope. Figure 13 illustrates the physical appearance of the smart glasses:

Figure 13: RealWear HMT-1 smart glasses (left) and their small screen (right)

Figure 14 provides a screenshot taken from the small screen of RealWear HMT-1:

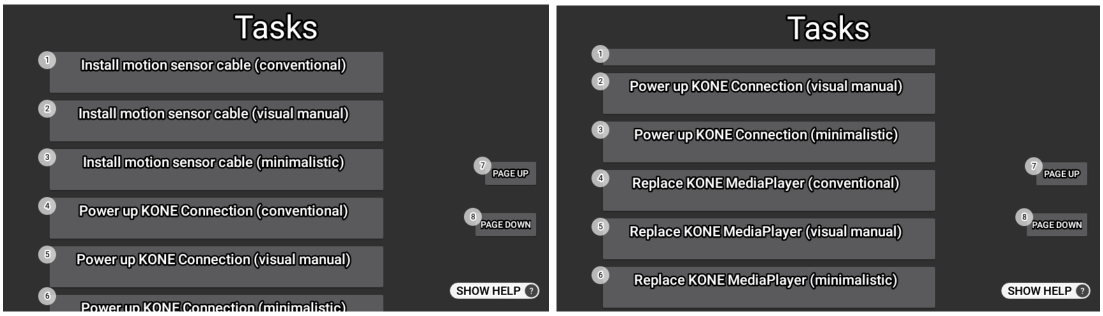


Figure 14: Task versions on RealWear HMT-1 menu

The materiality of this medium is a digital screen, which has certain properties: it has a 2D spatial extent, it is dynamic in the sense that content can be rendered on the screen again and again, and there is an interactive component to the medium, which takes place using the following multiple sensory modalities:

- Visual
  The glasses operate as a screen on top of one eye, still enabling the user to see the real world.
- Augmented reality (AR)
  The glasses have a camera to capture real world for AR rendering on top of the picture.
- Voice control
  The user interface operates with voice commands.
- Audio
  The internal speaker can play sounds for the user or the user can connect headphones to the glasses.

From these sensory modalities, I utilised the visual and voice control modalities in the design experiment. Later studies can address the remaining sensory modalities, for example, steps could be read aloud by RealWear HMT-1. Figure 15 illustrates the sensory modalities of RealWear HMT-1 that were utilised in my design experiment:

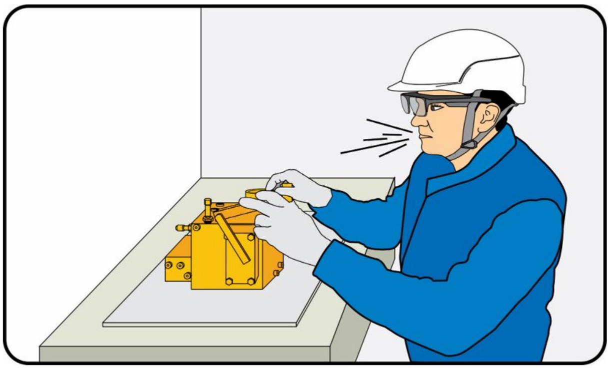


Figure 15: Visual and voice control sensory modalities in use for hands-free interaction between technician and smart glasses

In collaboration with my KONE colleagues and Realmax Oy, an application called Work Instructions was developed, to which we published the converted instructions in DITA XML format. The application development is out of scope of this thesis, but the implications of such development on professional technical communication are discussed in Chapter 5.1.

3.4 User test environment

The design experiment tasks “Install motion sensor cable” and “Power up KONE Connection” are performed on an elevator car roof, which is located inside the elevator shaft. A real elevator shaft is a dangerous environment, where only trained elevator techicians can work in. Together with my KONE colleagues, we simulated the shaft environment and the elevator car roof with a pallet that had a plywood fixed on the top. The mock-up car roof was placed safely into a meeting room. During the test, the test users stepped on top of the pallet as if they were working on a real car roof. Figure 16 illustrates an elevator car roof and how it was simulated in a meeting room:

Figure 16: Real elevator car roof in elevator shaft (left) and elevator car roof simulation with pallet in meeting room

The third task, "Replace KONE MediaPlayer", is performed in a relative safe public building lobby, not in the elevator environment. The task is similar to connecting and disconnecting cables with any media or content streaming device connected to a consumer TV. Therefore, the task could also be simulated in a meeting room, by using the meeting room monitor as a TV screen. Figure 17 illustrates how a pallet and plywood combination was used for a safe simulation of fixing KONE MediaPlayer close to the TV:

Figure 17: Building lobby with TV simulation in meeting room

To fine tune the test setup and converted instructions, a pre-test user performed the tasks and commented on the setup and tasks to validate the test environment before the real user test. The actual test users, many of them professional technicians, commented the test setup to be a good simulation of the real environment, apart from better lighting, cleaner surfaces and safer conditions.

4 Analysis

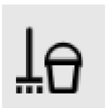
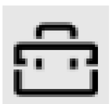
This chapter analyses the semiotic multimodality found from the data, and the actions performed when producing the conversions of the design experiment according to the theories of visual manual and minimalism heuristics. Section 4.1 starts by describing a mandatory preparatory stage for the visual manual version, creating a library of symbols for replacing verbs. Section 4.2 recognises the multimodal qualities of the material, and then walks us through the conversion of each step in the instructions (the material) in detail. This gives us an understanding, how the conversion changes the instructions, which then provides us the necessary groundwork for further multimodal analysis in Section 4.3. In Section 4.3, a representative sample of the conversion gets a deeper analysis according to the theories of multimodality, focusing on the changes in the semiotic meaning-making resulting from the conversion.

4.1 Preparatory stage: populating the inventory of symbols for visual manual

The largest modification, both visually and work-wise, during the conversion was to replace verbs with symbols according to the visual manual theory. Having a symbol library to start with is a must, because one cannot advance in the visual manual creation without it. Although a systematic symbol creation is out of scope of this thesis, I developed a method for it for the specific context of my design experiment. To build the initial symbol library, I analysed the selected task topics and established a mixed process according to Scurati et al. (2018) and what KONE already had in their symbol library. Scurati et al. (2018, p. 71) note that ISO-standard symbols are a good starting point for their familiarity, and new symbols would benefit to follow their style. Because KONE already had a library of existing task symbols, it was more beneficial to use those and mimic their style when new task symbols needed to be created. Therefore, for warning symbols, I used the ISO-symbols and for task symbols, I either used one from the KONE library or created a new one with a KONE illustrator, mimicking the KONE style for consistency inside KONE. The approach enables consistency benefits for KONE, as the new symbols can be reused in the KONE library.

Table 1 lists examples of the symbols used in the conversions:

Table 1: Example symbols used in the design experiment

Symbol	**Type of symbol**	**Meaning**
	ISO warning (from KONE library)	Risk of electric shock
	ISO mandatory action (from KONE library)	Wear eye protection
	Task (from KONE library)	Connect cable
 + 	Task symbol combination (created from KONE library symbols)	Finalise maintenance
	Task (created for this thesis)	Go to elevator car roof
	Task (created for this thesis)	Use cable tie

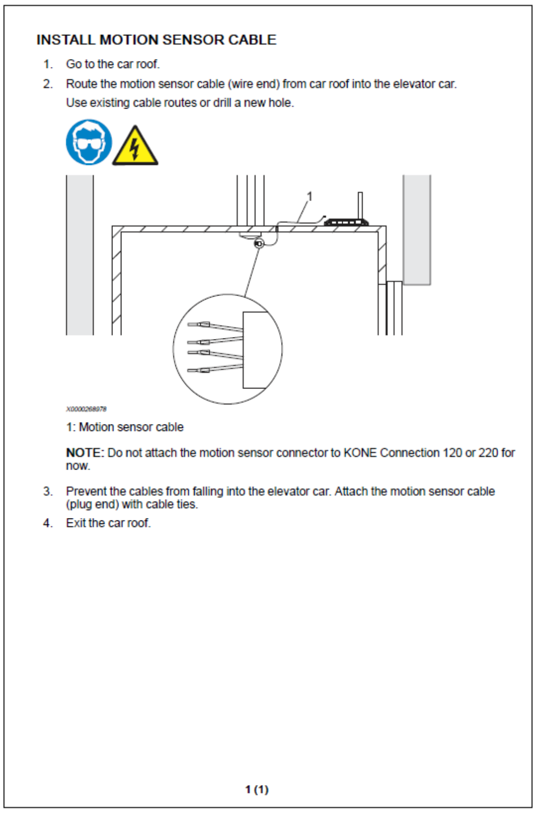
The understandability and guessability of the symbols are crucial (Wobbrock et al., 2005). As the shift from traditional documentation to visual manual is considerable, the symbols must be taught to the users beforehand (Gattullo et al., 2017, p. 1306; 2019, p. 285). The amount of authoring and training work needed for transitioning from one semiotic mode to another is further discussed in Sections 5.1 and 5.2.

4.2 Converting tasks for small screen

The following subsections illustrate the before and after appearance of each task, “Install motion sensor cable” (Section 4.2.1), “Power up KONE Connection” (Section 4.2.2) and “Replace KONE MediaPlayer” (4.2.3). Each section begins with a complete example taken from the original KONE PDF. I will identify the semiotic modes in operation on the PDF page, which is followed by screenshots and analytic description of the theoretically-informed conversions. For comparison, the *conventional* version illustrates the difference to the theoretically-informed redesigns of *visual manual* and *minimalism heuristics*. Due to the repetitiveness of the steps in the material, a repetitive analysis is fully described only once on its initial appearance and not repeated for similar steps afterwards.

4.2.1 Converting task “Install motion sensor cable”

This installation task describes installing a cable by routing it from the elevator car roof into the elevator car. For a more detailed description of the task, see Section 3.2. Figure 18 shows the original instructions:



INSTALL MOTION SENSOR CABLE

1. Go to the car roof.
2. Route the motion sensor cable (wire end) from car roof into the elevator car. Use existing cable routes or drill a new hole.

1: Motion sensor cable

NOTE: Do not attach the motion sensor connector to KONE Connection 120 or 220 for now.

3. Prevent the cables from falling into the elevator car. Attach the motion sensor cable (plug end) with cable ties.
4. Exit the car roof.

1 (1)

Figure 18: Original PDF of "Install motion sensor cable"

Figure 18 illustrates a semiotic mode that Bateman (2008, p. 175, 2011, pp. 25–27) refers to as text-flow. For more information on the typical modes operating on a document page, see Section 2.4.5. Hiippala (2016, p. 56) summarises the text flow as characterised by a one-dimensional structure that unfolds linearly, which in this case is emphasised by the steps. Text-flow does not exploit the layout space for making additional meanings. As also seen in Figure 18, Hiippala (2016, p. 56) states that it is common that text-flow is interrupted by (and integrated with) other semiotic modes. In Figure 18, text-flow is interrupted by the diagrammatic mode. The other modes are not bound by the one-dimensional linear structure, but the diagrammatic mode makes use of the two-dimensional extent of the page (Hiippala, 2016, p 56). Perhaps surprisingly, also the converted versions of the same material on the screen of the smart glasses (see, for example, Figure 20 or Figure 21) retain the text-flow mode. The difference is the absence of step numbers, and the addition of the navigational elements, for example, BACK and NEXT buttons, but the semiotic mode remains as text-flow. The same analysis holds true for the other material in my design experiment (“Power up KONE Connection” and “Replace KONE MediaPlayer”), as seen in Sections 4.2.2 and 4.2.3, so this analysis is not repeated for them separately.

Let us now focus on the small screen conversion actions that I performed according to the visual manual theory and minimalism heuristics.

Figure 19 illustrates the conversions of the title screen of “Install motion sensor cable”:

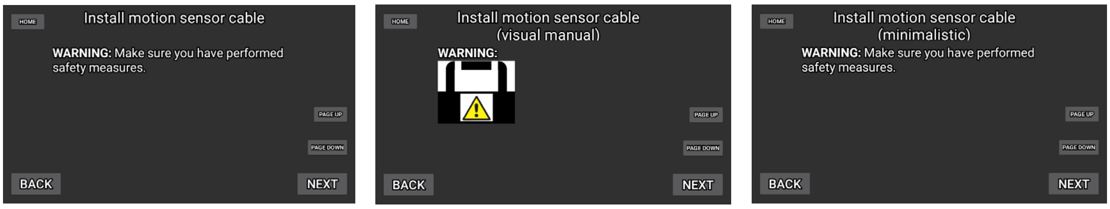


Figure 19 Title screen of “Install motion sensor cable”: conventional (left), visual manual (middle) and minimalistic (right)

The title screens of all conversions contain the heading, but in addition, I added a new warning, “Make sure you have performed safety measures”, that is not present in the original PDF for some reason. A moving elevator is dangerous, so certain safety measures to prevent accidental movement must be taken every time the elevator is maintained. The safety measures are trained standard actions that every qualified technician knows, so they do not need to be explained in detail. To represent implementing the safety measures in the visual manual version, a new symbol was developed, mimicking both the ISO warning sign and KONE style elevator symbol behind a safety fence. According to the visual manual theory, I chose not to add any text. For the minimalistic version, I used a textual warning typically used in KONE instructions. The WARNING text is automatically generated by the application.

Figure 20 illustrates the conversions of step 1 of “Install motion sensor cable”:

Figure 20 Step 1 of "Install motion sensor cable": conventional (left), visual manual (middle) and minimalistic (right)

Going to an elevator car roof is a standard installation and maintenance procedure that every trained technician knows how to do safely. For the visual manual version, I replaced all text with a new symbol with the same semantic content. Because KONE has an existing library of KONE-specific symbols, the new symbol was created mimicking KONE style, not ISO style. The new symbol is a combination of the KONE style elevator symbol, a person symbol, and an arrow pointing to the car roof.

Here, the visual manual version gets interesting when considering multimodal meaning-making. The visual manual version in Figure 20 (also already in Figure 19, but Figure 20 is a clearer example) requires a major shift in semiotic work required from the users. As seen with the analysis of the original PDF in Figure 18, users at KONE are accustomed to extracting the meaning of steps by following linear text-flow that is sometimes interrupted by the diagrammatic mode or an occasional symbol. Suddenly, this familiar text flow is replaced by a symbolic sign. According to discourse semantics, users have to contextualise symbolic signs in relation to what they are doing (Bateman, Wildfeuer and Hiippala, 2017, p. 116). In the visual manual version in Figure 20, there is no text anymore to work in combination with the symbol that would help the users to confirm whether their interpretation of the symbol is right or wrong. I will return to discuss how the shift to using more symbolic signs affected the test users in Section 5.2.

For the minimalistic version, I created a simple textual command in imperative mood, following the standards of technical writing. The result is similar to the conventional version.

Figure 21 illustrates the conversions of step 2 of "Install motion sensor cable":

Figure 21 Step 2 of "Install motion sensor cable": conventional (left, two screens due to scrolling), visual manual (middle) and minimalistic (right)

For the visual manual and minimalistic versions, I removed most of the step text, because the same semantic content is restated using the diagrammatic mode. According to the visual manual theory, I replaced the step verb with a "route wire" symbol that was created from scratch, mimicking KONE style. There are many cables on elevator car roof, so the cable to be routed must be named. Therefore, I left the cable name as text. To keep the text as short as possible, I chose not to use full sentences, since the "route wire" symbol already included the verb of the step.

For the multimodal meaning-making, this means that the users cannot seek much support for their interpretation of the diagram from the text, as the interplay of the text and the diagram is limited to just naming the cable. A similar shift in meaning-making repeats in many conversions to the visual manual, so I will not repeat the same analysis for the rest of the material.

For the minimalistic version, I followed the third principle of minimalism (support error recognition and recovery). I made the routing and drilling information as note to prevent users wasting time to think how to put the cable through the roof. According to heuristic 4.1 (be brief, do not spell out everything), I was able to shorten the step sentence by, for example, omitting cable connection and specific route description from the text, because the same semantic content is restated using the diagrammatic mode.

Figure 22 illustrates the conversions of step 3 of "Install motion sensor cable":

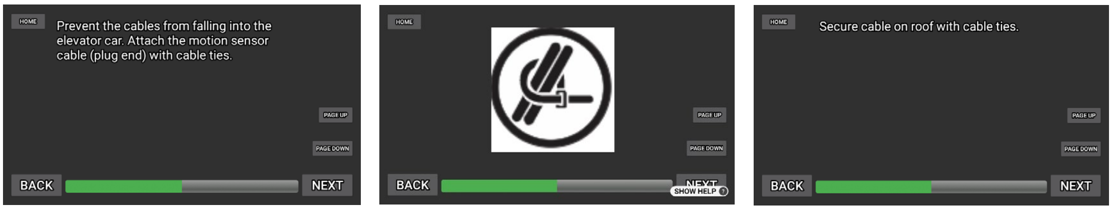


Figure 22 Step 3 of "Install motion sensor cable": conventional (left), visual manual (middle) and minimalistic (right)

The original text describes why the cable must be secured, but for a trained technician, this is self- obvious based on their experience. Therefore, I omitted the reasoning from both the visual manual and minimalistic versions, and for the visual manual version, this enabled replacing all text with a corresponding new symbol that was developed from scratch. Again, extracting the meaning of the single symbolic sign has the same challenges as described with Figure 20 already.

For the minimalistic version, I followed the minimalism heuristic 4.1 (be brief, do not spell out everything). I omitted much of the descriptive text. The original step itself is according to heuristic 3.1 (prevent mistakes whenever possible) and one could even argue that the reasoning for securing the cable could be kept in the step, and it would be according to heuristic 3.2 (provide error information when actions are error prone or when correction is difficult). However, due to the self-obvious nature of the task, I chose to omit it for small screen readability according to heuristic 4.1, and not to give explanations for the sake of explanations. This decision process shows that choosing the correct action is not always easy, but the author sometimes must rely on professional judgement, and knowing the readers' needs and expectations beforehand. But eventually, only the readers can evaluate if the result was understandable or not.

Figure 23 illustrates the conversions of step 4 of "Install motion sensor cable":

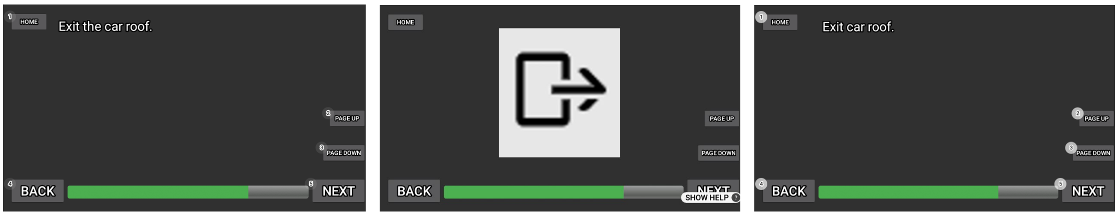


Figure 23 Step 4 of "Install motion sensor cable": conventional (left), visual manual (middle) and minimalistic (right)

Exiting the elevator car roof is another standard procedure that a trained technician is expected to know. Therefore, for the visual manual and minimalistic version, I utilised the same conversion process as seen in Figure 20 (Go to car roof).

Figure 24 illustrates the conversions of the end screen of "Install motion sensor cable":

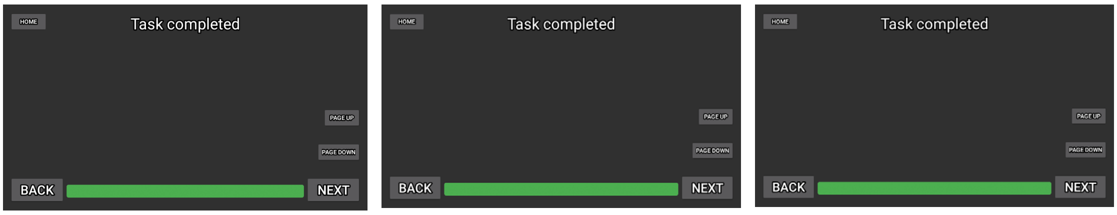
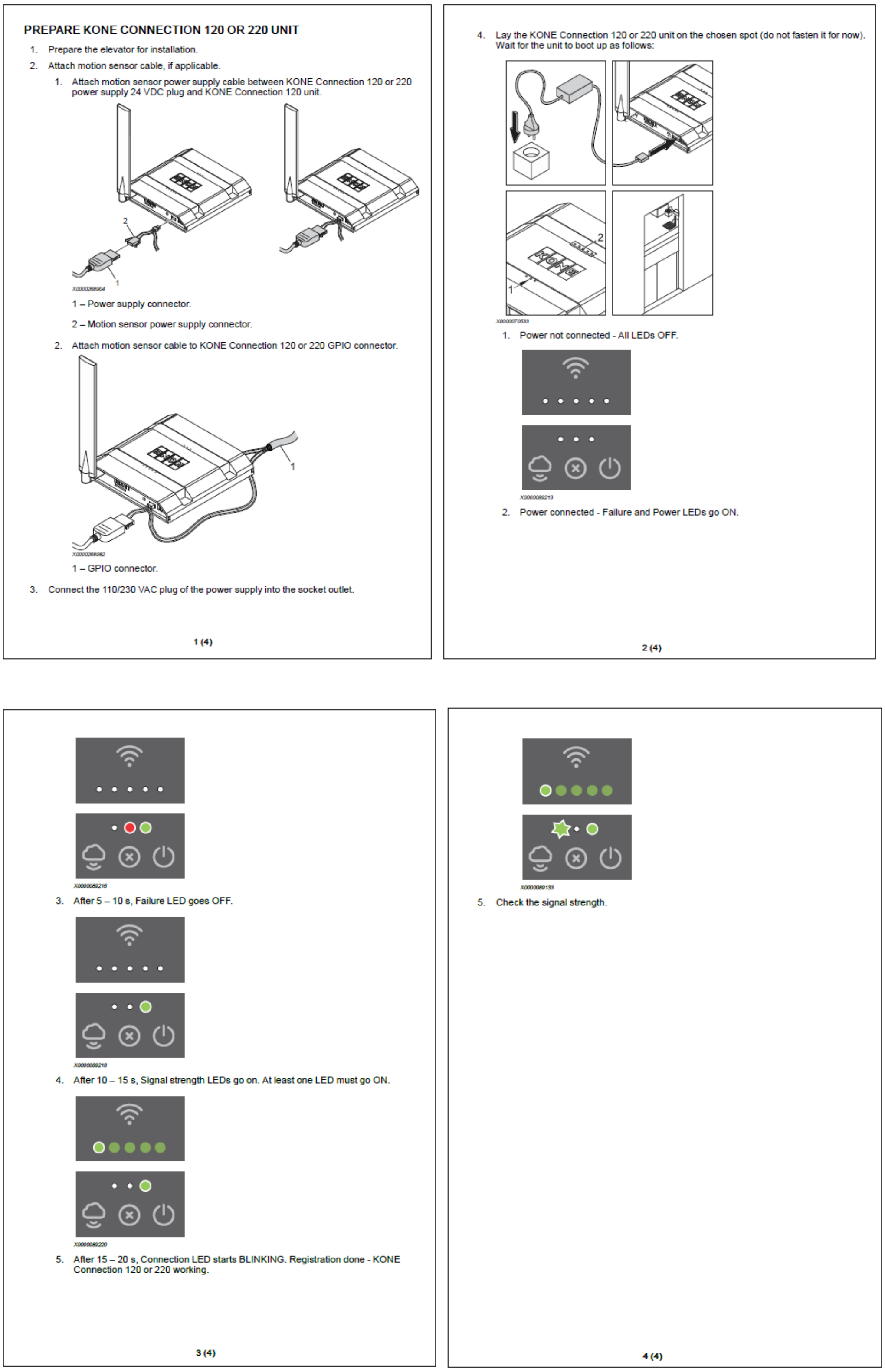


Figure 24 End screen of all tasks: conventional (left), visual manual (middle) and minimalistic (right)

To signal the user the end of the task, the application presented automatically generated text "Task completed". The "Task completed" screen is an additional element compared to the original PDF version, and it was added to all converted tasks. Because it is always the same, I will not repeat describing it for the other tasks.

4.2.2 Converting task "Power up KONE Connection"

This installation task describes connecting all required cables to KONE Connection unit and switching it on. For a more detailed description of the task, see Section 3.2. Figure 25 shows the original instructions:



PREPARE KONE CONNECTION 120 OR 220 UNIT

1. Prepare the elevator for installation.
2. Attach motion sensor cable, if applicable.
   1. Attach motion sensor power supply cable between KONE Connection 120 or 220 power supply 24 VDC plug and KONE Connection 120 unit.

      1 – Power supply connector.

      2 – Motion sensor power supply connector.
   2. Attach motion sensor cable to KONE Connection 120 or 220 GPIO connector.

      1 – GPIO connector.
3. Connect the 110/230 VAC plug of the power supply into the socket outlet.

1 (4)

4. Lay the KONE Connection 120 or 220 unit on the chosen spot (do not fasten it for now). Wait for the unit to boot up as follows:
   1. Power not connected - All LEDs OFF.
   2. Power connected - Failure and Power LEDs go ON.

2 (4)

3. After 5 – 10 s, Failure LED goes OFF.
4. After 10 – 15 s, Signal strength LEDs go on. At least one LED must go ON.
5. After 15 – 20 s, Connection LED starts BLINKING. Registration done - KONE Connection 120 or 220 working.

3 (4)

5. Check the signal strength.

4 (4)

Figure 25: Original PDF of "Power up KONE Connection"

Figure 26 illustrates the conversions of the title screen of "Power up KONE Connection":

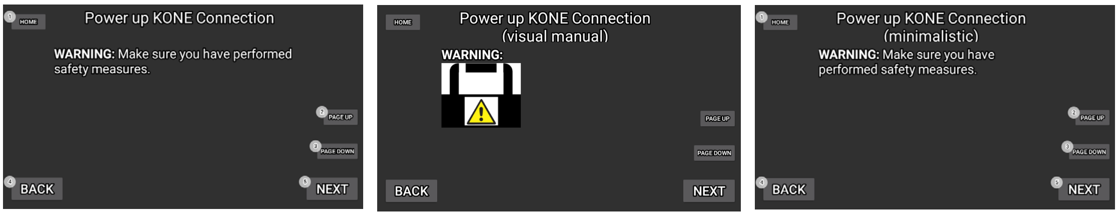


Figure 26 Title screen of "Power up KONE Connection": conventional (left), visual manual (middle) and minimalistic (right)

For the title screen, I utilised the same conversion process as seen in Figure 19 (title screen of "Install motion sensor cable").

Figure 27 illustrates the conversions of step 1 of "Power up KONE Connection":

Figure 27 Step 1 of "Power up KONE Connection": conventional (left), visual manual (middle) and minimalistic (right)

For going to the elevator car roof, I utilised the same conversion process as seen in Figure 20 (Go to car roof).

Figure 28 illustrates the conversions of step 2 of “Power up KONE Connection”:

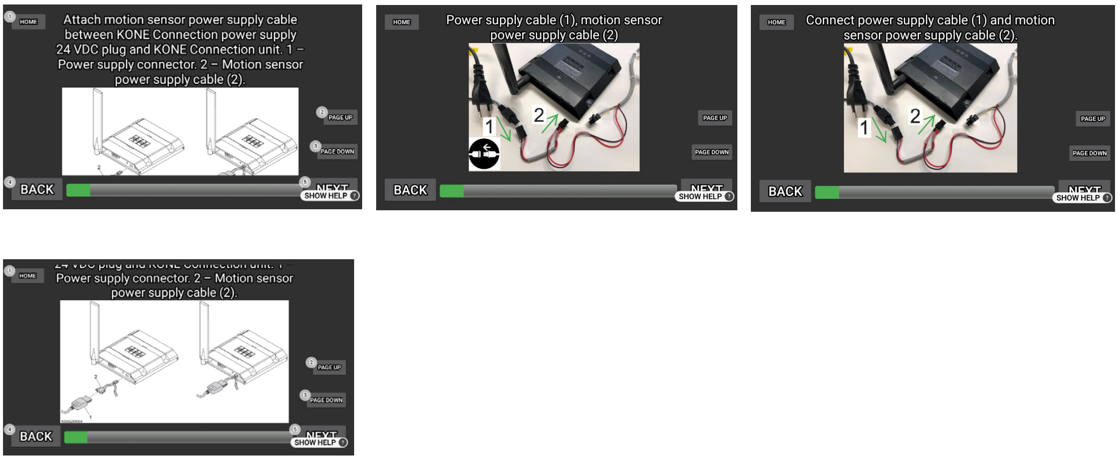


Figure 28 Step 2 of “Power up KONE Connection”: conventional (left, two screens due to scrolling), visual manual (middle) and minimalistic (right)

For the visual manual version, I again removed most of the step text, because the same semantic content is restated using the diagrammatic mode. My pre-test user commented that the line drawn diagram visible in the conventional version did not depict the cables and plug ends clearly enough on small screen. Therefore, I replaced the line drawn diagram with a realistic photograph with diagrammatic elements (callout numbers and respective arrows) drawn on top of it. The resulting new diagram is visually more similar to the actual task at hand. I replaced the step verb with a “connect plug” symbol that was available in KONE symbol library. For text, I applied the same strategy as seen in Figure 21 (routing the motion sensor cable).

For the minimalistic version, I followed the minimalism heuristic 4.1 (be brief, do not spell out everything). I was able to shorten the step sentence by, for example, omitting the device and detailed cable descriptions from the text, because again the same semantic content is restated using the diagrammatic mode.

Figure 29 illustrates the conversions of step 3 of “Power up KONE Connection”:

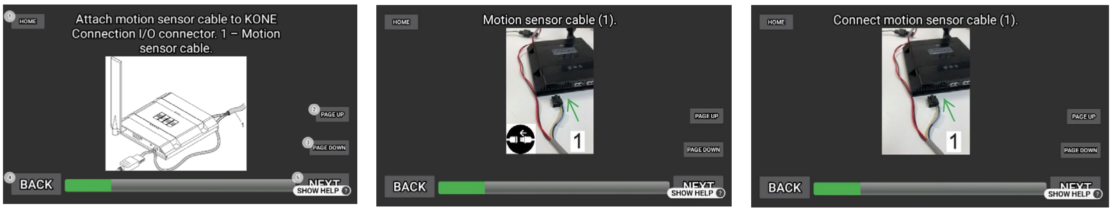


Figure 29 Step 3 of "Power up KONE Connection": conventional (left), visual manual (middle) and minimalistic (right)

Step 3 is about connecting another cable, so it is similar to step 2 (see Figure 28). I utilised exactly the same conversion process.

Figure 30 illustrates the conversions of step 4 of "Power up KONE Connection":

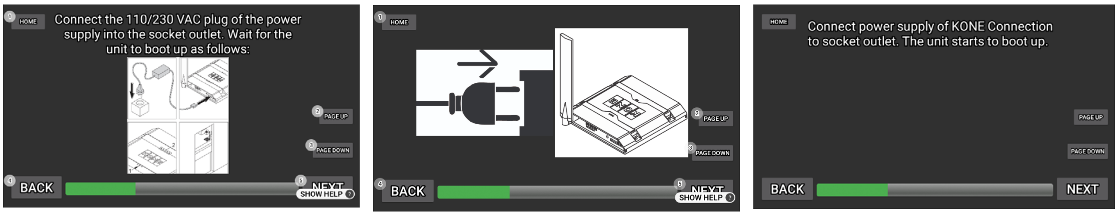


Figure 30 Step 4 of "Power up KONE Connection": conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I again removed all the step text, because the same semantic content is restated using a symbol and a line drawing of the device. I replaced the step verb with a "connect main plug" symbol that was developed from scratch. Even though all text was removed, this visual manual step was better understood in the user test than the steps where a single symbol appeared alone. To extract the meaning here, the symbol receives support from the picture of KONE Connection. When these two modes interact together, it helps the users to understand that it is the KONE Connection whose power cable must be plugged in.

For the minimalistic version, I applied heuristic 4.1 (be brief, do not spell out everything) similarly to the previous steps. In addition, and according to heuristic 3.3 (provide error information that supports detection, diagnosis, and recovery), I added "The unit starts to boot up" as an indicator for the user to progress to the next steps 5–9, where the boot up LED sequence that the user should monitor gets displayed.

Figure 31 illustrates the conversions of steps 5–9 of “Power up KONE Connection”:

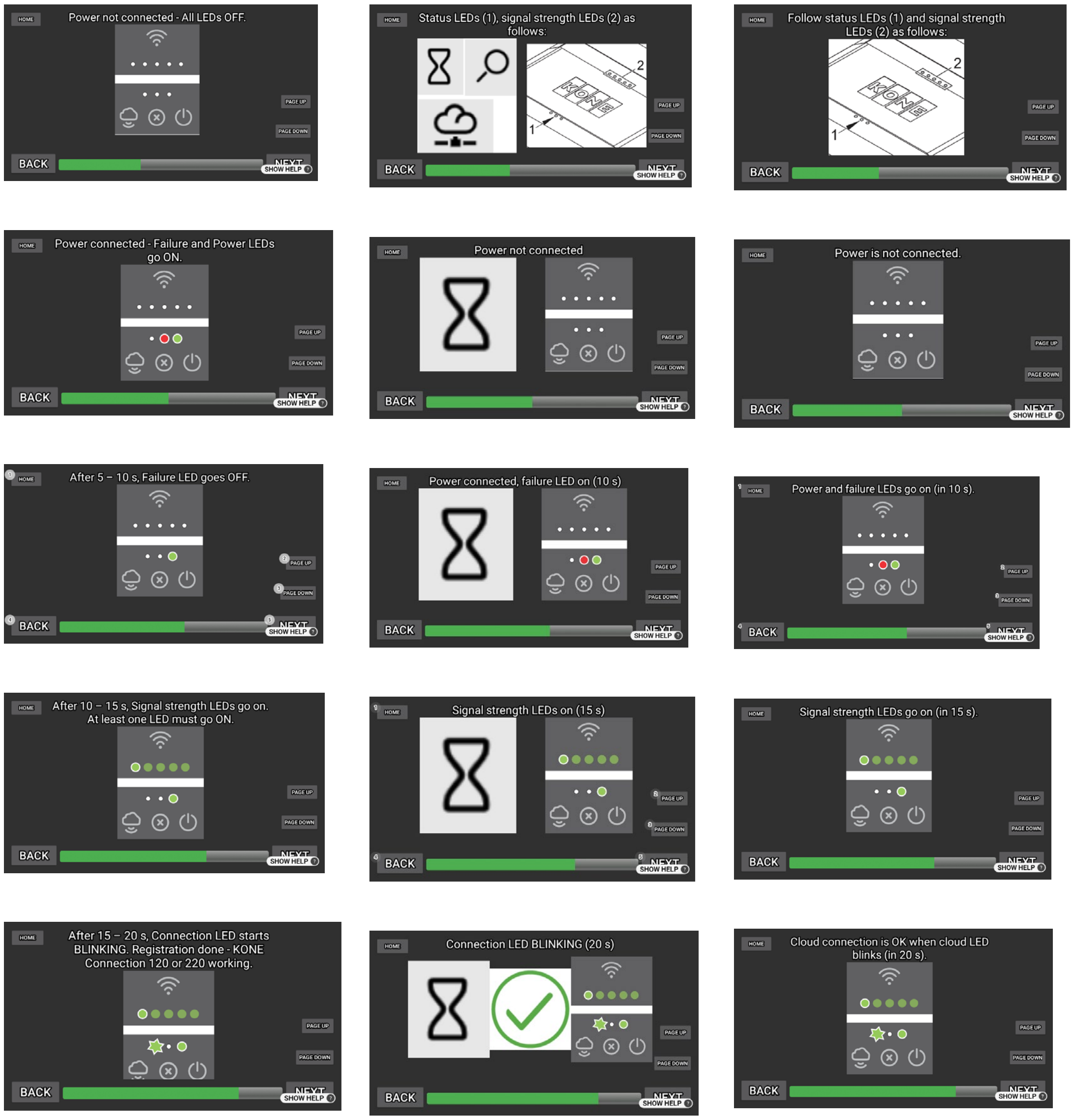


Figure 31 Steps 5 – 9 of “Power up KONE Connection”: conventional (left), visual manual (middle) and minimalistic (right)

After connecting the main power, the user follows up if the unit boots up and establishes a connection. In the original PDF, the boot up sequence was presented with text and images of how the UI LEDs on the unit represent the phases of the boot up sequence. For the visual manual and minimalistic versions, I kept the words “as follows” to signal the user that the following steps must be checked to understand the boot up sequence as a whole. For the visual manual version, I replaced the verbs with symbols according to the visual manual theory, but preserved a short textual explanation of what each LED means. As the meaning of the LEDs is the main point of these steps, I felt that text was the only semantic mode to carry the meaning precisely. In step 5 of the visual manual (the first screen of Figure 31, in the middle), there is a further shift in meaning-making: the user must deduct the required step action from the combination of three symbols: the hourglass, the magnifying glass and the cloud. According to Bateman, Wildfeuer and Hiippala (2017, p. 115), such symbol clusters are read grammatically and not lexically. The meaning of the symbol cluster builds up after reading them together, like a sentence. I will return to sign clusters when analysing them with theories of multimodality in Section 4.3.4.

For the minimalistic version, I omitted the "check" verbs from the individual LED indication steps because of Heuristic 4.1 (be brief, do not spell out everything) and preserved a short textual explanation of what the LEDs mean.

Figure 32 illustrates the conversions of step 10 of "Power up KONE Connection":

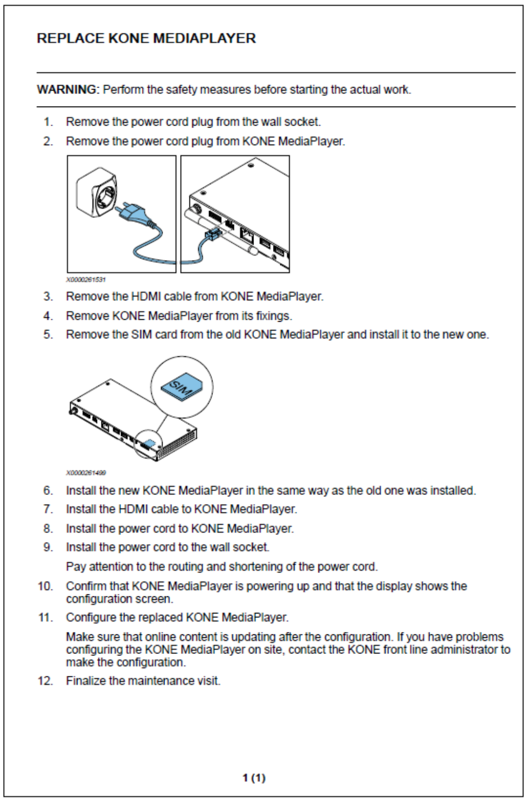
Figure 32 Step 10 of "Power up KONE Connection": conventional (left), visual manual (middle) and minimalistic (right)

As the last step, the user checks that the signal strength is sufficient for the unit to operate. For the visual manual version, I replaced the verb with an existing symbol from KONE library, but added the object of the check (signal strength LED) as text. Again, I felt that text was the only semantic mode that could carry the meaning precisely.

For the minimalistic version, I also added the object of the check as text. This is according to the third principle of minimalism (support error recognition and recovery).

4.2.3 Converting task "Replace KONE MediaPlayer"

This maintenance task describes replacing a broken KONE MediaPlayer unit. For a more detailed description of the task, see Section 3.2. Figure 33 shows the original instructions:



REPLACE KONE MEDIAPLAYER

WARNING: Perform the safety measures before starting the actual work.

1. Remove the power cord plug from the wall socket.
2. Remove the power cord plug from KONE MediaPlayer.
3. Remove the HDMI cable from KONE MediaPlayer.
4. Remove KONE MediaPlayer from its fixings.
5. Remove the SIM card from the old KONE MediaPlayer and install it to the new one.
6. Install the new KONE MediaPlayer in the same way as the old one was installed.
7. Install the HDMI cable to KONE MediaPlayer.
8. Install the power cord to KONE MediaPlayer.
9. Install the power cord to the wall socket.
   Pay attention to the routing and shortening of the power cord.
10. Confirm that KONE MediaPlayer is powering up and that the display shows the configuration screen.
11. Configure the replaced KONE MediaPlayer.
    Make sure that online content is updating after the configuration. If you have problems configuring the KONE MediaPlayer on site, contact the KONE front line administrator to make the configuration.
12. Finalize the maintenance visit.

1 (1)

Figure 33 Original PDF of "Replace KONE MediaPlayer"

Figure 34 illustrates the conversions of the title screen of "Replace KONE MediaPlayer":

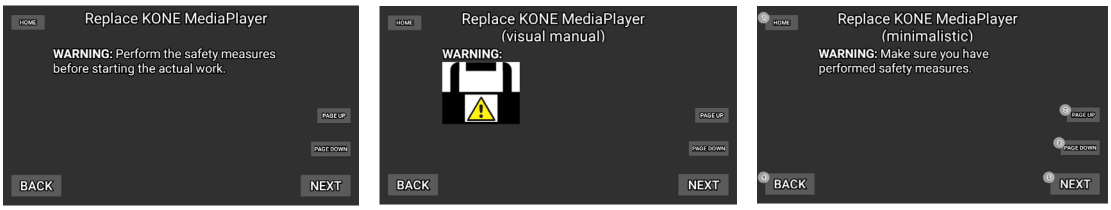


Figure 34 Title screen of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

Again, for the title screen, I utilised the same conversion process as seen already in Figure 19 (title screen of “Install motion sensor cable”).

Figure 35 illustrates the conversions of step 1 of “Replace KONE MediaPlayer”:

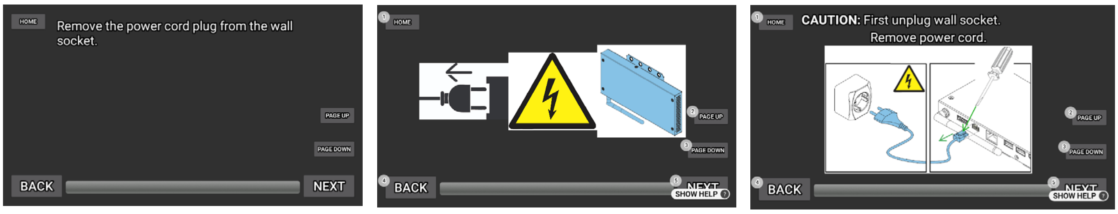


Figure 35 Step 1 of “Replace KONE MediaPlayer”: conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I removed all the step text, because the same semantic content is restated using symbols and a line drawing of the device. There are many cables behind a TV in the lobby, but this time, I was able to avoid naming the cable with text by adding a picture of the device whose power cord must be disconnected. Additionally, I used the electric shock ISO symbol to underline that it is the power cord that must be disconnected. Naturally, it also signifies the danger of electric shock. Deducting the meaning is similar to Figure 30 (action symbol working in combination of a picture of the operated device), with the addition of the warning symbol.

Because both ends of the power cord must be removed, I combined the two first steps of the original into one step for the minimalistic version. According to the third principle of minimalism (support error recognition and recovery), I made the wall socket removal as note to make the step safer and to avoid equipment damage. The order of removal was present in the original, but as two different steps.

Figure 36 illustrates the conversions of step 2 of “Replace KONE MediaPlayer”:

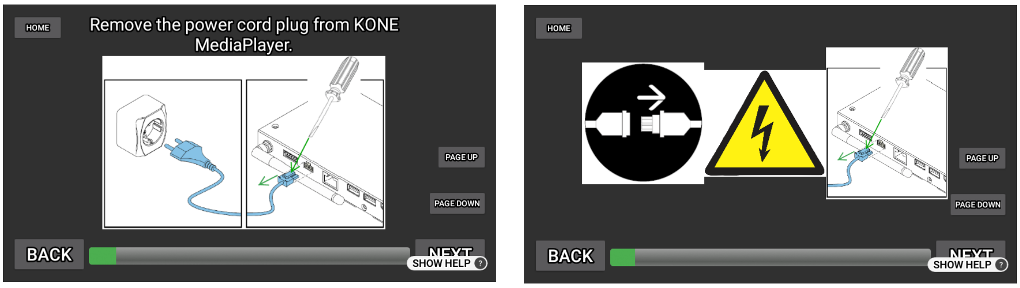


Figure 36 Step 2 of "Replace KONE MediaPlayer": conventional (left) and visual manual (right) (minimalistic step 2 was combined with step 1)

Step 2 is similar to step 1, so I used the same conversion methods for the visual manual version as described above for step 1.

Figure 37 illustrates the conversions of step 3 of "Replace KONE MediaPlayer":

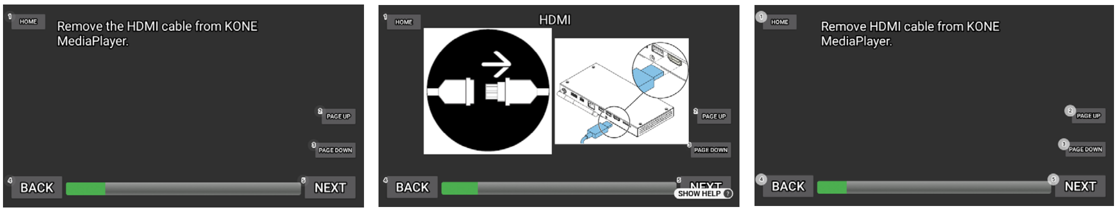


Figure 37 Step 3 of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I utilised the same methods as already described for earlier steps (see, for example, Figure 21 and its description).

For the minimalistic version, one must consider that the HDMI cable is such a common household cable nowadays that no illustration is necessary. I simply restated the text from the conventional version.

Figure 38 illustrates the conversions of step 4 of "Replace KONE MediaPlayer":

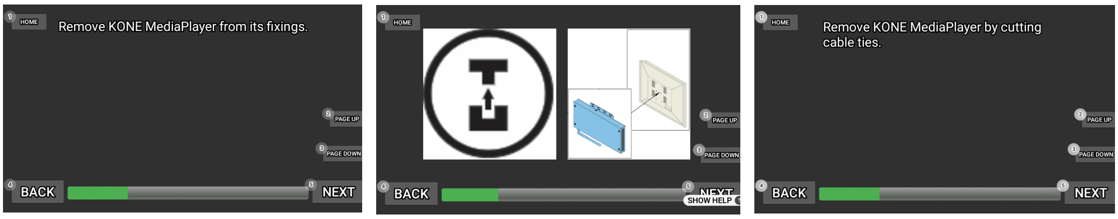


Figure 38 Step 4 of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I utilised similar methods as earlier, and replaced the step verb with a "remove component" symbol.

For the minimalistic version, one must consider that the device is fixed behind a TV with cable ties (easy to recognise), so again, no illustration is necessary. Even though it is common knowledge that cable ties must be cut to be removed, I included this to the minimalistic text to save the users' time to think about it. This is loosely according to the first principle of minimalism (choose an action-oriented approach) and heuristic 3.1 (prevent mistakes whenever possible).

Figure 39 illustrates the conversions of step 5 of "Replace KONE MediaPlayer":

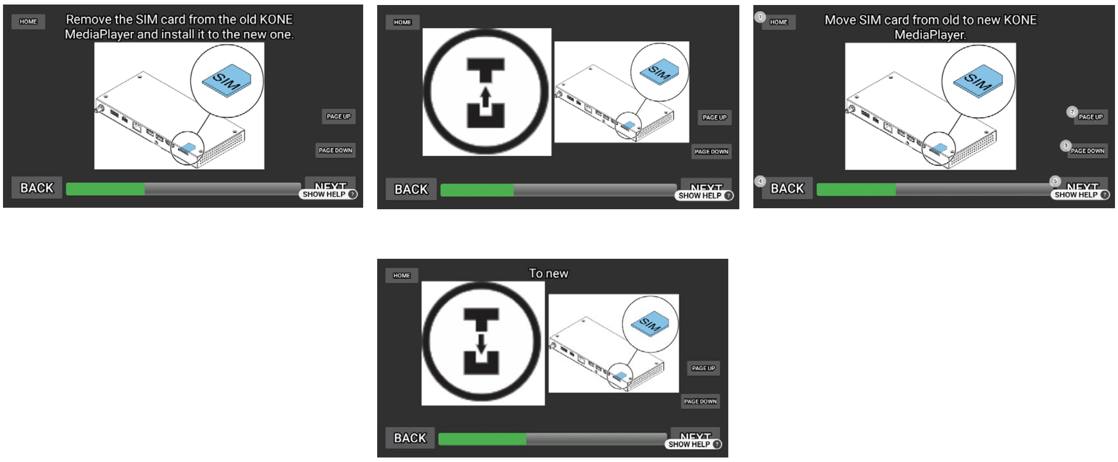


Figure 39 Step 5 of "Replace KONE MediaPlayer": conventional (left), visual manual in two steps (middle) and minimalistic (right)

The SIM card is not broken, so it must be moved from the broken device to the new one. For the visual manual version, I used the "remove component" and "install component" symbols to replace the verbs, and the main image identified that it is the SIM card that we are moving. However, I felt that I needed to add the text "To new" to make sure that the user does not put the SIM card back to the broken device. In the original manual, moving the SIM card was one step, but based on the pre-test user's comments that suddenly there are two actions in one step, and the second one can easily get omitted, I added a step for installing the SIM to the new unit.

However, for the minimalistic version, to support heuristic 1.2 (encourage and support exploration and innovation) and heuristic 4.1 (be brief, do not spell out everything), I justified keeping the step as one step. Hence, I used the word "move" to signify that you must do more than just remove the SIM.

Figure 40 illustrates the conversions of step 6 of "Replace KONE MediaPlayer":

Figure 40 Step 6 of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I utilised similar methods as earlier. In the original and visual manual versions, the installation procedure was presented step-by-step. However, for the minimalistic version, to support heuristic 1.2 (encourage and support exploration and innovation) and heuristic 4.1 (be brief, do not spell out everything), I only stated that the installation must be performed in reverse order. Additionally, according to the third principle of minimalism (support error recognition and recovery), I made using cable ties as note to make the installation tidier and to avoid equipment damage caused by loose cables.

So far with the minimalistic step conversions, extracting the meaning of the steps has been quite similar to the conventional version, except for occasionally combining a couple of steps together or leaving out an obvious diagram, when the same message could be achieved with text alone. In Figure 40, the sequence of how the steps are presented makes a shift in the meaning-making: leaving out the next steps by only stating that the installation must be performed in reverse order to the removal steps forces the users to either deduct the reinstallation sequence based on their memory or experience, or navigating back the step screens to revisit the order of the removal steps. This change in the “flow” of the steps requires different strategies for extracting the meaning of the step: instead of deducting the meaning from the text-flow of a single screen, the whole communicative situation changes to using the voice-controlled, immutable ergodic navigational user interface of the smart glasses. I will return to this topic in Section 5.2, when discussing how the test users reacted to such changes in the step-flow.

Figure 41 illustrates the conversions of step 7 of “Replace KONE MediaPlayer”:

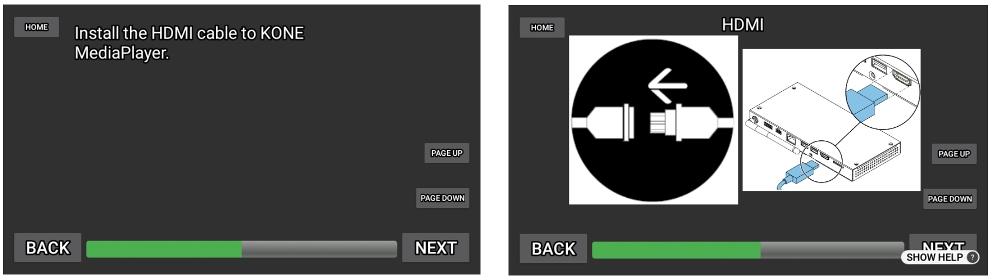


Figure 41 Step 7 of “Replace KONE MediaPlayer”: conventional (left) and visual manual (right) (minimalistic was combined with step 6)

For the visual manual version, I utilised similar methods as earlier. There is no minimalistic version as it was combined with step 6.

Figure 42 illustrates the conversions of step 8 of “Replace KONE MediaPlayer”:

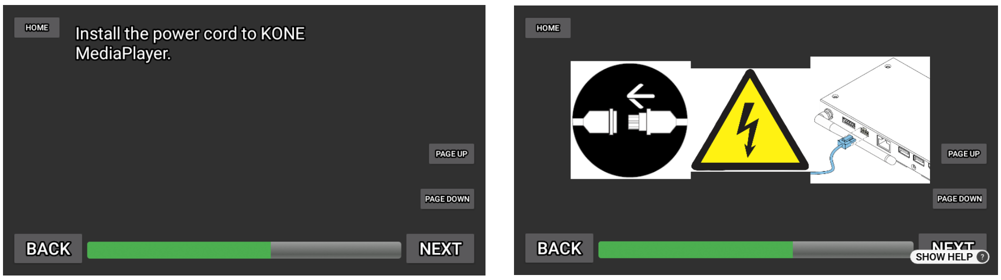


Figure 42 Step 8 of “Replace KONE MediaPlayer”: conventional (left) and visual manual (right) (minimalistic was combined with step 6)

For the visual manual version, I utilised similar methods as earlier. There is no minimalistic version as it was combined with step 6.

Figure 43 illustrates the conversions of step 9 of “Replace KONE MediaPlayer”:

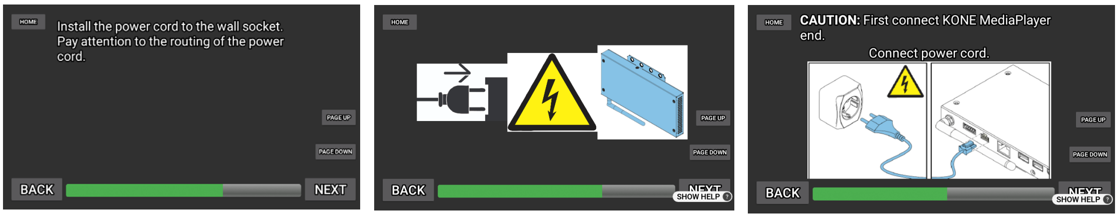


Figure 43 Step 9 of “Replace KONE MediaPlayer”: conventional (left), visual manual (middle) and minimalistic (right)

Step 9 includes reverse actions to step 1 (see Figure 35 about disconnecting the power cord and its description), so I utilised the same strategies.

Figure 44 illustrates the conversions of step 10 of “Replace KONE MediaPlayer”:

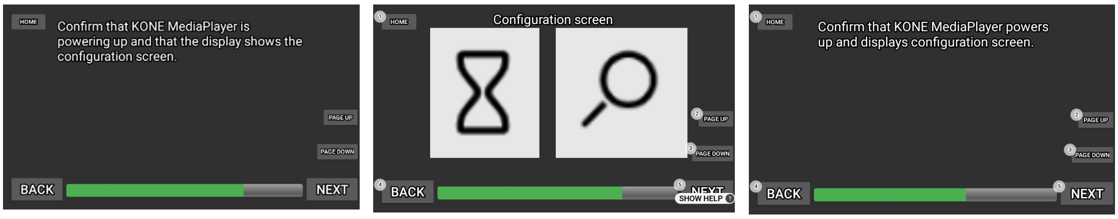


Figure 44 Step 10 of “Replace KONE MediaPlayer”: conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I removed most of the step text, except that I identified that the user must wait for the “configuration screen” to display on the TV. I replaced the step verbs with a sign cluster consisting of “wait” and “check” symbols.

For the minimalistic version, I utilised the same conversion process as earlier (see, for example, Figure 20 and its description).

Figure 45 illustrates the conversions of step 11 of “Replace KONE MediaPlayer”:

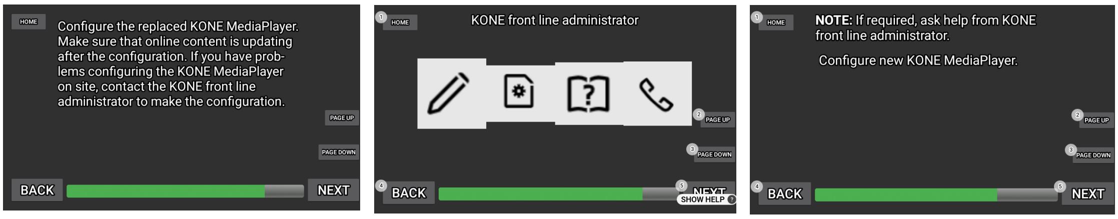


Figure 45 Step 11 of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

For the visual manual version, I again removed most of the step text, except that I identified who to call if problems arise. To experiment with the grammatical interpretation of sign clusters, I replaced the step verbs with "edit", "configure", "problems" and "call" symbols to create a four-part sign cluster to signify that the user must configure the new device, and a KONE front line administrator is able to help if problems. I will return to sign clusters when analysing them with theories of multimodality in Section 4.3.4.

For the minimalistic version, I shortened the text, and made asking for help as note to make it stand out according to the third principle of minimalism (support error recognition and recovery).

Figure 46 illustrates the conversions of step 12 of "Replace KONE MediaPlayer":

Figure 46 Step 12 of "Replace KONE MediaPlayer": conventional (left), visual manual (middle) and minimalistic (right)

For both versions, I utilised similar methods as earlier. For the visual manual version, the symbols "clean" and "maintenance" were used to replace the verbs as a sign cluster to signify that the user must clean up the workplace.

4.3 Multimodal analysis of conversions

After seeing the results of the conversions in the previous sections, let us revisit them with full focus on analysing them with theories of multimodality. I will structure my analysis according to the phases of the multimodal navigator suggested by Bateman, Wildfeuer and Hiippala (2017, pp. 211–232):

- Characterise the communicative situation
- Distinguish genre and medium
- Decompose canvases
- Analyse

4.3.1 Characterise the communicative situation

Before any analysis, one should decide in advance the aspects of the multimodal situation that are in the focus of the study. One must also understand that the object of analysis must be captured, and the more that the captured recording differs from the actual situation, there will be an increasing amount of loss in the data (Bateman, Wildfeuer and Hiippala, 2017, p. 212). To limit the scope of the multimodal analysis, I set out to record only the instruction screens as screenshots and some photos of the spatial working situation, the elevator shaft and building lobby. As this thesis focuses on the conversions to small screens, I will limit the focus of my multimodal analysis mostly to the canvas on which the instructions are presented. This means that the loss in my recorded data is the micro-ergodic navigational properties of the HMT-1 smart glasses, the Work Instructions application, and the spatiotemporal situation of performing the tasks on the elevator car roof and building lobby. However, if the spatiotemporal dimension of the work environment clearly has an influence on the multimodal meaning-making, I will also note those observations. After all, the goal is to facilitate the interpretation of the instructions for the technician.

As mentioned already in Chapter 4, I chose the step 5 of “Replace KONE MediaPlayer” task for my full multimodal analysis. As illustrated by Figure 47, the step is one of the clearest examples of what each conversion theory has changed when compared to the conventional version: verbs were replaced with symbols in the visual manual version, and text was clarified in the minimalistic version:

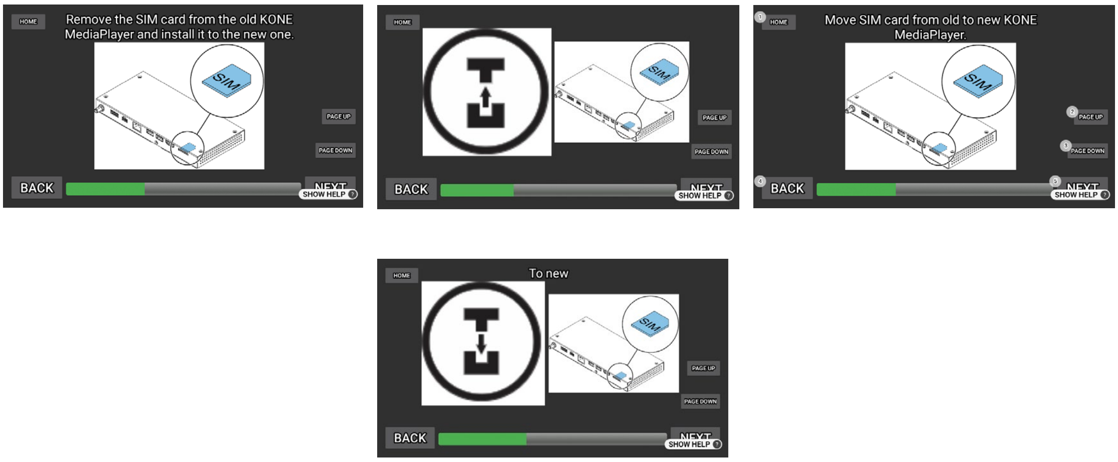


Figure 47 Object of multimodal analysis: steps for moving SIM card

Figure 48 illustrates the spatiotemporal context of the “move SIM card” step:

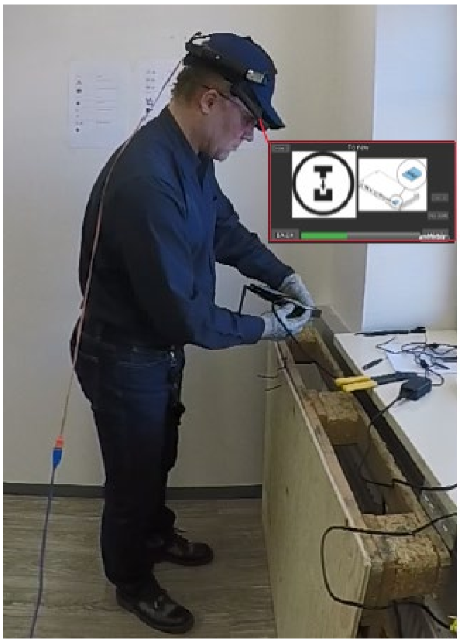


Figure 48: Work context of “move SIM card” step: KONE MediaPlayer replacement in simulated building lobby

4.3.2 Distinguish genre and medium

Bateman, Wildfeuer and Hiippala (2017, p. 129) define genre as a pattern of conventions or communicative activities developed by a society or culture for performing communication in some situation. As illustrated by Figure 49, Bateman, Wildfeuer and Hiippala (2017, p. 228) propose the following decision procedure to identify the properties of the canvas under focus:

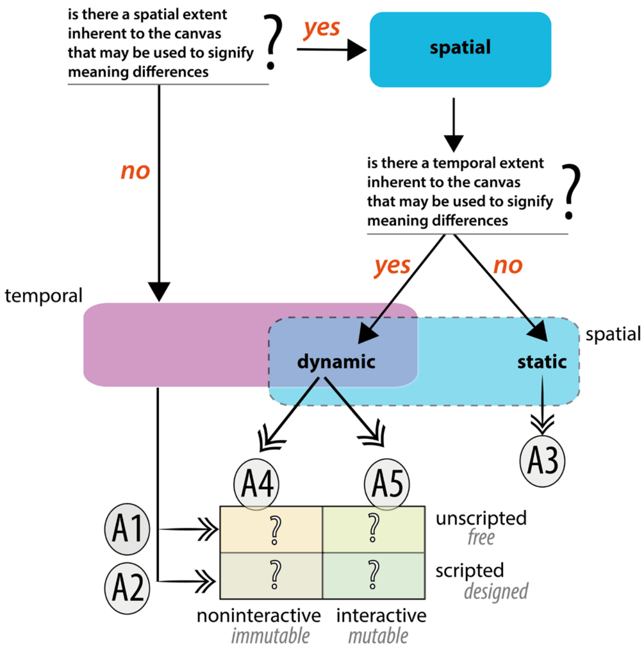


Figure 49: Decision procedure for working from the canvas of any medium to the particular use case areas (Bateman, Wildfeuer and Hiippala, 2017, p. 228)

According to Bateman, Wildfeuer and Hiippala (2017, p. 229), identifying the properties will help limiting the high-level focus to the most significant of these various possibilities:

A1 = Temporal unscripted

A2 = Temporal scripted

A3 = Spatial static

A4 = Spatial dynamic

A5 = Spatiotemporal interactive

Analysing the whole multimodal situation depicted in Figure 48 according to Figure 49 reveals the complexity of using instructions in context. There is clearly the spatial dimension of the technician working in the lobby, which classifies as spatiotemporal and interactive (A5). But when we focus on the technician following the instructions on the screen of the smart glasses, the result is different: temporal and scripted (A2). And to continue, the subcanvas (the screen), on which the actual diagrammatic instructions are, is spatial and static (A3).

According to Bateman, Wildfeuer and Hiippala (2017, p. 287), diagrammatic instructions represent the visual procedural instruction genre. There are more than diagrams in my material, such as text and symbols, but diagrams are clearly the dominant semiotic mode. At this point, we can at least safely state that the main genres in my material are assembly and repair instructions, which mainly belong to the visual procedural instruction genre. However, Bateman, Wildfeuer and Hiippala (2017, p. 288) note that we must be constantly aware that sometimes a particular genre, such as instructions, can draw on resources that are not typically associated with it. For example, when a "call for help" -comic strip appears inside an IKEA assembly instruction.

4.3.3 Decomposing canvases

Before analysis, we must understand how the multimodal situation at hand is composed, especially what the physical medium and its canvases are, and most importantly, how the canvases differ from each other in terms of their multimodal affordances. That is why we must make a decomposition of canvases to understand what kind of analysis is relevant for each canvas. Bateman, Wildfeuer and Hiippala (2017, p. 213) state that sorting out the media and decomposing its canvases to subcanvases and slices brings organisation to complex situations, and saves a lot of work and confusion, because a medium can have nested or even overlapping virtual canvases where activities happen. Bateman, Wildfeuer and Hiippala (2017, p. 219) further state that even the subslices of canvases can be further segmented. As the result of the decomposing and slicing, we arrive to the meanings carried by semiotic artefacts familiar from other disciplines.

Figure 50 decomposes the multimodal situation in the "move SIM card" step.

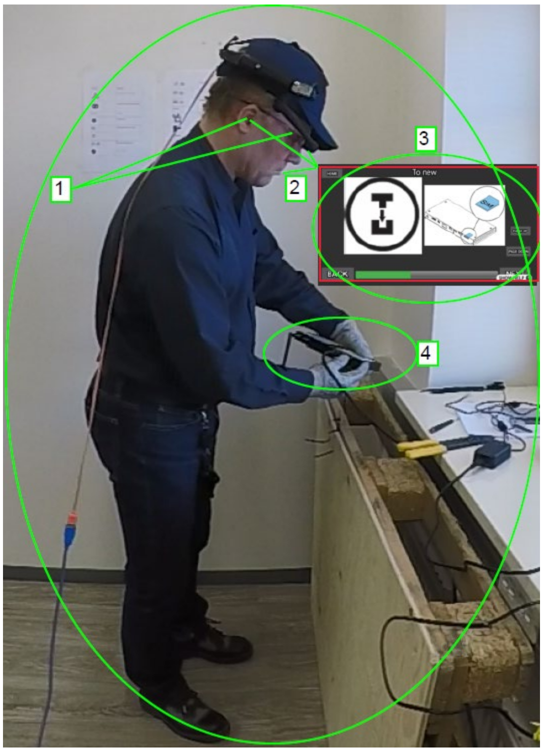


Figure 50: 'Move SIM card' step canvases decomposed

The most inclusive canvas is the 3D mutable ergodic lobby where the technician is working (1). Even though the HMT-1 glasses (2) are a medium, they can be labelled as an immutable ergodic subcanvas in the lobby situation, with the individual step screens (3) as subcanvas slices inside the glasses. The step screens are 2D static, observer-based and micro-ergodic. The final subcanvas is the 3D mutable ergodic equipment (4) that the technician is touching. From each of these canvas slices, we can already understand that different modes can be employed in them. For example, the touch employed on the replaced KONE MediaPlayer cannot be used with the onscreen instruction, but instead, voice commands are possible. Similarly, occasional passers-by can start a conversation with the technician, but the technician is not able to conduct a similar mutable ergodic discussion with the glasses nor the instruction screens.

To simplify, Figure 51 illustrates taking the canvases as slices according to Bateman, Wildfeuer and Hiippala (2017, p. 218):

Figure 51: "Move SIM card" step canvas slices

Bateman, Wildfeuer and Hiippala (2017, p. 221) note that we must limit the amount of canvases to analyse to serve the research question. The slicing above serves this purpose. Because the focus of this thesis is converting traditional instructions on small screens, I will limit my canvas analysis to the step screens. However, the whole 3D mutable ergodic lobby situation can have an influence on how the message in the instructions carries over to the technician, for example, noise, lighting conditions, other people, even differences in the actual equipment as compared to what is depicted in the instructions can have an influence. This can be addressed in further research.

With an understanding to limit the analysed material, I will now work forward to analyse the "move SIM card" step screen in more detail.

4.3.4 Analysis

When looking at the big picture, all my data is a typical instruction, consisting of text and pictures, utilising linear and micro-ergodic text-flow, image-flow and page-flow. In all my data, the images are mainly diagrams made with line drawings or photographs, and symbols. One could expect to find also information graphics, but there are none in my selection of data.

I will now set out to see how the used conversion theory changes the multimodal meaning-making. The starting point and baseline for the analysis is the conventional step as illustrated in Figure 52:

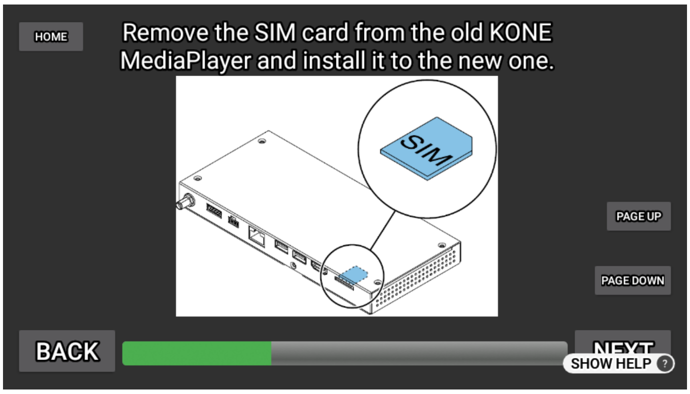


Figure 52: Conventional version of "move SIM card" step

Even though the instruction is on digital screen, showing one step at a time, one step screen still constitutes a two-dimensional, static canvas working with a page metaphor. In other words, the instructions are presented in a two-dimensional layout space simulating a page. The communicative situation of interacting with the medium is immutable and ergodic: the user can go back and forth through navigation, but cannot manipulate the content presented. Instead of showing many steps on one A4 paper page, we now see one step on a small screen at a time. Still, the page layout remains the same between steps. On the page, we have text and a diagram. As a result, there is the diagrammatic semiotic mode in action with page composition. There is also media convergence at play, where another media depicts another media, for example, think about iPad depicting a newspaper (Bateman, Wildfeuer and Hiippala, 2017, p. 225). Here, the smart glasses are depicting photos and diagrams. There could even be video or animations, or other types of media depicted on the smart glasses. So actually, the conversions are drawing minimally on the potential of the medium.

Bezemer and Kress (2016, p. 81) note that unlike, for example, a video, diagrams can only hold "frozen" actions, that is, they cannot show how the action unfolds in time. Yet, this is not a problem for understanding the step, because after the reader has identified the semiotic mode of the picture as the diagrammatic mode, instead of, for example, a painting, the expectations are adjusted accordingly for understanding the meaning of the whole step, including the diagram (Bateman, Wildfeuer and Hiippala, 2017, p. 288). Bateman, Wildfeuer and Hiippala (2017, p. 61) discuss the term "abduction" (reasoning to the best explanation) by Peirce, to be crucial for semiotic modes. In abduction, the interpreter interprets sign differently than with logical deduction. In logical deduction, the interpreter arrives to the *correct* answer, but with abduction, the interpretation can also go wrong. In the diagram, we see the device itself and two circles with a line between them highlighting the SIM card. Because we have understood that the diagrammatic mode is in action, the most plausible explanation that may be recovered through abduction is that there are not two SIM cards in the picture, but the bigger one is a zoom out of the smaller one, and the smaller one is actually indicating the correct location of the same SIM card. I will return to abduction with sign clusters shortly.

As Bezemer and Kress (2016, p. 7) state, modes appear in combination, for example, as a mixture of image, writing and page layout, constituting *multimodal ensembles*. Regarding the conventional version, it is important to note that the text could be used without the diagram to carry over the meaning of the step, but the diagram alone is not able to achieve the same meaning, since there is no action depicted. The circles and line of the diagram only highlight the location of the SIM card, thus participating to the meaning-making as a multimodal ensemble. It is typical in technical communication that the interplay between text and images works well together (Schriver, 1997, p. 408). They get the meaning signified better than either one alone (Virtaluoto, Suojanen and Isohella, 2021, p. 23).

Next, Figure 53 shows the minimalistic version:

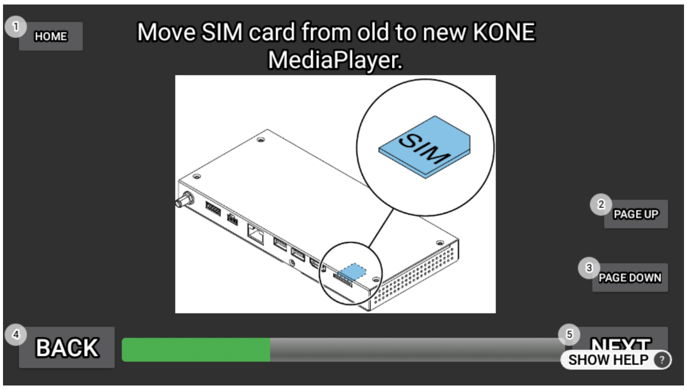


Figure 53: Minimalist version of "move SIM card step"

Again, we have the diagrammatic semiotic mode in action with page composition, similar to the conventional version. The difference is in text, which is shorter, but the main difference is that the verb usage has changed. Instead of using two different verbs, “remove” and “install”, the minimalistic version has replaced them with just one verb, “move”. In interplay with the same diagram as in the conventional version, there is a difference in the meaning-making, which was also found during the user test: if the technician is used to or expecting that there will be a separate step for installing the SIM card to the new device, the meaning of “move” can be neglected. Some testers paused at this point, wondering what should be done with the SIM card in their hand, until they either deducted it from their experience with SIM cards, or reread the step and understood the meaning of “move”. In the steps before this step, there had been only one action per step, for example, “remove HDMI cable”. Even though “move” is linguistically one action, it requires at least three actions in this maintenance case. First, you remove the SIM from the broken device, then you take a new device in your hand, and lastly, you install the SIM to the new device. Apart from this, the multimodal ensemble and its meaning-making operation is similar than in the conventional version.

The result that there is not much different multimodally on the screen of the smart glasses when compared to the conventional version, or even to the traditional PDF instructions, may sound striking at first. However, Bateman, Wildfeuer and Hiippala (2017, p. 348) confirm this by stating that the information offering of the new digital media is not automatically different than in traditional media, and that we should not take any assumptions as the starting point. New media often uses the same multimodal means as traditional media. Again, think about newspaper views modelled on iPads.

Lastly, Figure 54 illustrates the visual manual version:

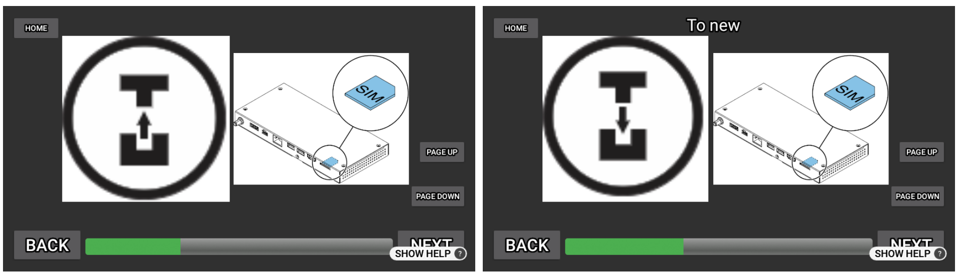


Figure 54: Visual manual version of the “move SIM card” step as two steps

Again, we have the diagrammatic semiotic mode in action with page composition, similar to the other versions, but in addition, the theory has added symbolic signs and reduced text considerably. The page-flow is now different, because the task has been split into two steps. This step was easily comprehended in the user tests because of the clear page-flow, and there is only one symbol per step here. Bateman, Wildfeuer and Hiippala (2017, p. 115) discuss the lexical versus grammatical meaning of signs. If symbols appear in clusters, their meaning becomes grammatical, and if they appear alone, the meaning is lexical. In the current step, the meaning is lexical, but on some instances in my data, the symbols must be read grammatically. This is important to note, because the current step was fast to read and easy to comprehend, whereas with grammatical sign clusters as illustrated by Figure 55, the test users spend considerably more time reading the symbols. After reading, many were still not sure of the overall meaning. This interpretation is supported by Scurati et al. (2018, p. 69), when they state that the number of symbols presented at the same time should be minimised for optimising user performance.

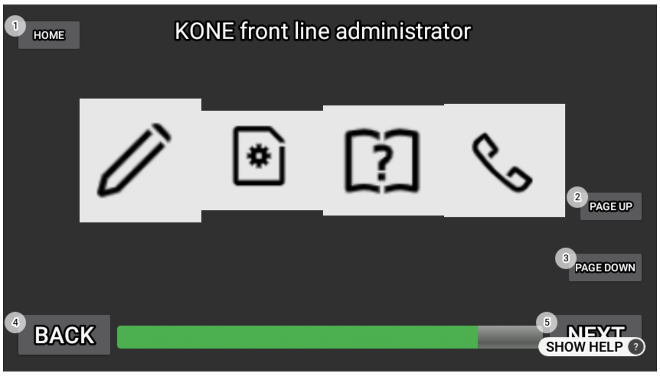


Figure 55: Sign cluster in step 11 of "Replace KONE MediaPlayer"

As far as the actual symbols go, Bateman, Wildfeuer and Hiippala (2017, p. 60) state that interpreting the meaning of a symbolic sign is fixed by convention, for example, the word "dog" does not resemble a dog, but by convention, we know what "dog" means. After a standardisation of the symbol set used according to the visual manual theory, an initial training is required for easy operation with symbols (Gattullo et al., 2019, p. 285). However, evaluating how instruction users understand the meaning of symbols is out of scope of this thesis, but it is an area that should be studied more. Most of the test users highlighted the meaning of knowing the symbols in advance, even though the symbols were presented to them in advance.

What the users of instructions must do with symbols is related to discourse semantics: they have to contextualise the symbol in relation to what they are doing, for example, moving the SIM card. Discourse semantics means making the symbol (sign) interpretable in context. Actually, all semiotic modes have discourse semantics, otherwise we would not be able to relate their interpretations in context (Bateman, Wildfeuer and Hiippala, 2017, p. 117). In Figure 54, the readers contextualise the generic ‘remove component’ and ‘install component’ symbols with the provided diagram of the SIM card location to arrive to the meaning that it is the SIM card that they must move to the new device. Yet, we can see problems with symbols in some steps in the material. A good example is illustrated in Figure 55 with the sign cluster. In the user tests, there were some signs of abduction going wrong regarding the step in Figure 55. Some users understood that they must write something down, even though the pen symbol means “edit”, and the goal of the step is to configure the new device into operation, and if problems, call the front-line administrator for help. So, in visual manual, wrong interpretations with abduction are a risk, when users are trying to interpret the symbols, especially sign clusters.

5 Discussion

This chapter places the analysis conducted in Chapter 4 to the context of a daily life of a technical communication professional. Section 5.1 starts with practical, subjective observations on how the researched conversion and design theories would affect authoring such instructions. Although the full results of the user study are not in scope of this thesis, Section 5.2 pulls in key observations from how the test users understood the new ways of meaning-making that resulted from the conversions. Again, I will utilise theories of multimodal communication to systematically understand some of the key observations. Based on all these observations, this chapter concludes with ideating implications for future research benefiting Industry 4.0 instructions in Section 5.3.

5.1 Implications on instruction production

This section discusses the wider implications of small screen authoring than what was conducted in Chapter 4. The focus is on the technical communication business side, on the possible effects on a company's instruction production.

5.1.1 Content management system and application development

An obvious but often neglected area of digital authoring is that one must have the full pipeline from authoring to publication. It is clear that one must have the physical device, for example, a mobile phone, or in the scope of this thesis, the HMT-1 smart glasses. What is often forgotten is that the content does not magically appear on the device in a usable and user-friendly way. The company needs to acquire a new or develop their existing content management system (CMS), with a dedicated publishing pipeline and application support in the form of stylesheets that render the source content used by the technical writer, for example, DITA XML, onto the screen of the output device. On the market, there are platforms to create and show content on small screens, but few are tailored for industrial instruction production without company-specific tailoring. The tailoring can require extensive development by a company, especially for small screen publishing. And for a company to embark on the tailoring of such a CMS, as stated earlier, there is a growing need for guidelines about how to work with the new technological devices, but few exists (Tham et al., 2018, p. 179).

Regarding the actual work required to enable small screen publishing, the usability, layout and how one must work with the source XML content so that it displays in the application in the optimal way is not a trivial task and calls for further research. In the scope of this thesis, a KONE usability specialist was used to comment on the first iterations of the Work Instructions application on the smart glasses. For example, the things we worked with included which picture formats and sizes to use, and how to display the XML elements, since the space on the small screen is limited. Since pure XML is free of any layout or font definitions, the stylesheet that produces the layout for the screen had to be developed by a person capable of programming stylesheets. Lastly, the usability specialist commented the logic of the screen progression, marginals, used font sizes, positioning of the images and text as well as the navigational properties of the application. Arriving to the layout and user experience seen in Section 4.2 required multiple iterations. Figure 56 illustrates an example of issues that required fine tuning:

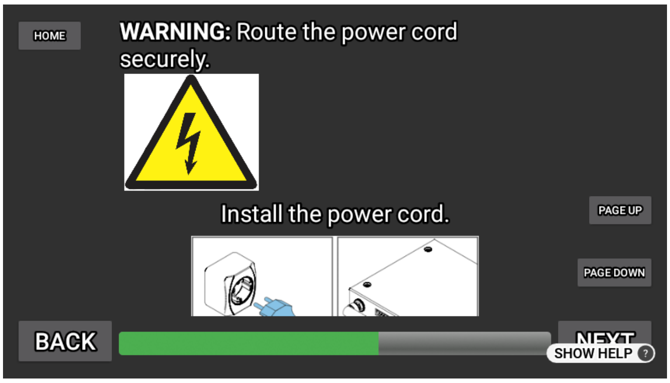


Figure 56: Example application development layout with image positioning and image size issues

As noted by Siltanen and Heinonen (2020, p. 107), this thesis also reinforces the call for more research to optimise the creation and user experience with Industry 4.0 instructions.

5.1.2 Instruction authoring

Whether we want it or not, the business implications of doing anything differently or in the worst case, in addition to the existing processes must be considered. Many technical products change quickly, so the respective technical communication must be easy to update. Furthermore, translations are often required, so it is crucial to consider the translation effort whenever we change the source content. Reducing text in favour of more digital, visual manuals can result in translation cost reduction in the long run, but industrial companies still have a need of paper manuals. Therefore, a compromise should be made to convert the existing paper-format documentation into a digital format, with the eventual goal to go towards AR interfaces (Gattullo et al., 2017, p. 1304). As an example, KONE has hundreds of thousands of pages of existing documentation, ranging from design, installation, maintenance and user manuals to sales guides. And given the global setting of any such company, not all front-line operations can take new digital technology in use simultaneously. So, the paper manual is there as a backup, and in many cases, it is also the main platform to deliver instructions. For these reasons, it is not an option to abandon the current technical communication production processes, but all new development should be added as an option. Yet, the use of a *separate* system for small screen production would "double" the production costs. Here, standards such as DITA XML are on top of the game, where you can author the same content once in the same production system, and publish it to multiple output channels, for example, as traditional PDF for paper printing and as HTML for new digital channels.

With the business requirements as the background, I made some observations what it was like to produce the visual manual and minimalistic conversions. As noted earlier, I made the conversions from existing instructions of KONE Oyj, and with the same authoring tools (DITA XML). Only the publication channel was different, as discussed above. For visual manual, once the symbol library was fully created and available, it felt fast, natural and easy to evaluate what is the main verb in a step and select the respective symbol and immediately insert it to the task. The subjective feeling was that the writing process felt faster and easier than traditional textual writing, because you do not have to consider your choice of words, you only pick a symbol and you are done. The workload shifted from authoring the textual content to authoring the main image of the step. The main image had to interplay with the chosen action symbol, so that without text, the symbol and the main image fully created the meaning of the step. What happened was that the meaning-making power shifted from the text to the multimodal ensemble of the main image and the symbol of the step. As noted earlier, modes appear in combination, for example, as a mixture of image, writing and page layout (Bezemer and Kress, 2016, p. 7).

Regarding the minimalistic conversion, there was less impact on authoring, because the existing KONE technical communication style guide has its roots in minimalism, whether knowingly or just by sharing global best practices between other companies. As a result, many minimalism heuristics appear in the KONE style guide, although they are just not called as such. However, the design experiment in this thesis showed that in addition to “slashing the verbiage” most commonly associated with minimalism (Redish, 1998, p. 219), the authors need to become more aware of what *not* to include in an instruction. This is related to the first principle of minimalism heuristics, “choose an action-oriented approach”, and heuristic 4.1, “be brief, do not spell out everything” (van der Meij and Carroll, 1998, p. 21). At KONE Oyj, content production is a collaboration between many people. Several input providers, commonly known as subject matter experts (SMEs), give information from *their* perspective to the instruction. If the authoring team fails to assess the minimalism heuristic 1.3, “respect the integrity of the user’s activity” (van der Meij and Carroll, 1998, p. 21), thus keeping *all* provided information in the instruction, the result can be a long instruction. From such an instruction, the readers cannot find the part that *they* would need. At this point, the required minimalist action by the author would be to know what the readers need, and remove everything else from the instruction, hence fulfilling the minimalistic design idea of the instruction. Sometimes subject matter experts, especially safety-concerned ones, feel that everything *they* know, or can imagine happening with the task, must be documented. To apply minimalism effectively, also subject matter experts, in addition to the technical writer, must change their traditional ways of working (Virtaluoto, Suojanen and Isohella, 2021, p. 10). For example, they must get used to providing “incomplete information” (Rosenbaum, 1998, p. 119).

5.2 Implications on instruction readers

Based on the conducted user tests, this section discusses what small screen instructions would require from the readers, and more importantly, what requirements the readers have for such instructions. After all, instructions are there for the reader, not the other way around. With Industry 4.0, the small screen becomes the central piece of displaying just the right amount of technical information. Arguably, even when technology evolves, for example, enabling hologram screens or similar technologies of any size, the instructions still benefit from digital compactness. Yet, there are certain limitations and preferences readers are either accustomed to or expecting, as noted while conducting the user tests. These we must not forget. For example, scrolling on a small screen should be avoided or at least minimised (Churchill and Hedberg, 2008, p. 883), and the design of information should take the context of the task at hand into account to help the readers to make meaning of complex information (Albers, 2002, p. 51). Users prefer to see the context in which a single task is performed to avoid hesitance. They like to know where they are coming from and where they are going to in regards of the overall continuum of the work they are conducting. In a traditional paper-format manual, the table of contents is typically the key factor in showing the context. A similar type of table of context should be secured to small screen instructions, a design that was not included in this thesis, but clearly something to be addressed in further research. During the user test preparation, the pre-test user commented that it was difficult to get started without knowing the context, so for the actual user test, this was remedied by carefully talking through the context before the test users started reading the instructions. As a result, none of the test users had difficulties with the context. But in real life, there will be nobody to give you the context, so it must be embedded to the instructions or clarified in some other way.

What we can also see is that readers must get accustomed to more diverse multimodal meaning-making. How much more depends on many things discussed in this thesis, for example, what are the multimodal affordances of the used canvas, the used small screen design theory, and how much development a company is willing and capable of investing business-wise. Potentially, there can be a considerable shift between media, which affects the role of the instruction user, moving from being a sign consumer of a static and immutable page-flow document to a micro-ergodic participant in a spatiotemporal augmented reality environment, or even more. When interacting with instructions on printed paper or rendered on a static PDF document, the user takes the role of an observer, a follower of a static and permanent object that makes use of the 2D representation space, the layout. The same holds true to some extent to the design experiment in this thesis, which did not venture beyond to, for example, a smart watch 2D screen design. However, further studies should take more advantage from the emerging potential in augmented reality, where the user role can be changed into an interactive participant in a 3D spatiotemporal environment.

Regarding the conversions in this thesis, it was already noted that the visual manual requires teaching the readers the symbols beforehand, but there were further observations regarding multimodal meaning-making with the “more symbols less text -approach”. I started out with the idea that with the most obvious symbols, for example, with “plug in” and “plug out” a cord, the symbol alone would be enough to carry the meaning over to the reader. But already the pre-test user noticed that although you know which component you are working with, there might still be other devices with their cords nearby. In my design experiment, the observation resulted always including the main image to each step, where the main image depicted the component to work on. Only the symbols without any risk of confusing to which component they apply to could be used alone, for example, “take elevator out of use”, “go to elevator car roof” and “finalise maintenance”. Yet with these, there is a possibility of confusion if one is working with an elevator group, that is, several elevators side-by-side. These types of observations become clear only when testing something with real equipment, not necessarily when you are drafting a piece of instruction in the isolation of your office desk.

At the other end of the spectrum, the test users commented that using many picture modes, diagrams, photographs or pictorial symbols, in the same step, is not good either. The use of too many picture modes is illustrated in Figure 57:

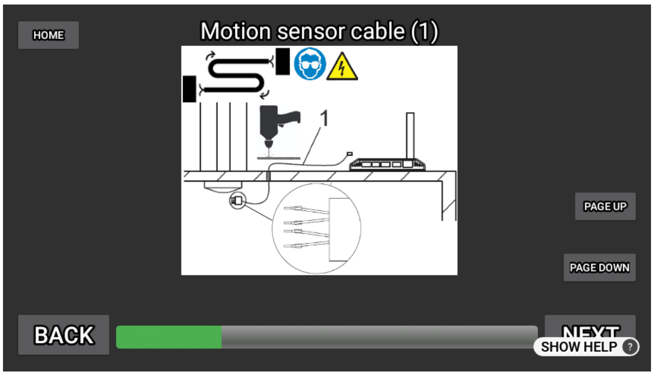


Figure 57: Routing a cable through elevator car roof in step 2 of "Install motion sensor cable" task, visual manual version

Figure 57 presents a simple action: install a cable through the car roof into the elevator car, to the space where the passengers are. Yet, we have one action symbol for routing the cable, a mandatory action symbol to wear safety goggles, a warning symbol for electricity, another action symbol to use a drill, and the main picture in the form of a diagram. Lastly, the name of the cable to be routed is indicated with text, a callout number (1) links the text and the diagram, and a zoom bubble indicates which end of the cable must go inside the elevator car. Here, users commented that it is hard to know where to start reading the diagram, and to distinguish the main action of the step. Some test users also commented reading meticulously from left to right, so the items located to the right were in greatest danger to be missed. In general, the visual manual steps that were understood the best were steps with one action symbol and one main image. Even adding a warning symbol to the step slowed down the overall meaning-making for the user. Thus, we can say that the best practice seems to strike a balance that is tilted more towards too little visual modes rather than too many. Many users ideated that a short animation would work better instead of a combination of still image modes for the meaning-making of a complicated step.

The amount of information given in one step is equally important also for the meaning-making in the minimalistic version, according to the test users. In the minimalistic version for moving the SIM card to the new KONE MediaPlayer, the removal and installation was bundled into the same step by saying "Move SIM card to new device" (see Figure 53), whereas in the visual manual version (see Figure 54), there are two separate steps for removing and installing. In the user tests, some users neglected reinstalling the SIM card with the minimalistic version, and the SIM card was momentarily left out from the new device, until the users realised their mistake at a later stage. There were no such problems with the visual manual version. According to the test users, it is better to limit actions per step as much as possible, and if the previous steps clearly have a "rhythm" of one simple action per step, changing the "rhythm" can confuse extracting the meaning of the step.

At the other end of the spectrum, some experienced users commented that for simple tasks, for example, connecting a commercial power or HDMI cable, it can also be a surprise that there actually are detailed steps presented. Such simple steps also disturbed the meaning-making and the flow of some of the users' performance. This is related to the minimalism heuristic 1.3, "respect the integrity of the user's activity". We should know the users well enough to present only the information they need, not what they already know of (van der Meij and Carroll, 1998, p. 21). In the user test, this reflected so that some experienced users conducted phases of a task, and when checking back to the instructions, they realised that they had performed some of the steps already.

Lastly, Figure 58 illustrates how difficult it is to get several pictorial modes working together, or even a simple diagram to be understood equally by all users (abduction going wrong):

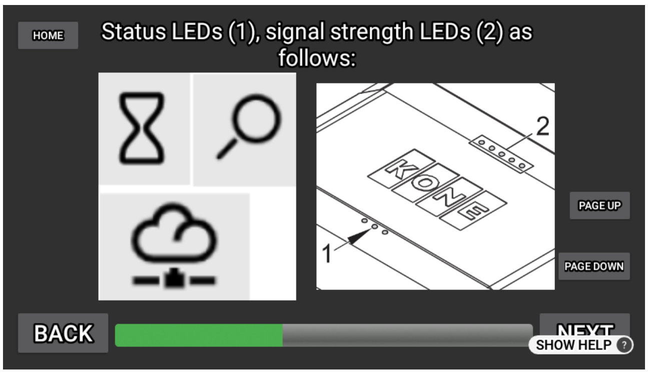


Figure 58: Step 5 of "Power up KONE Connection' task", visual manual version

In the diagram on the right, the arrowhead is pointing to a group of three LEDs, with the intended meaning that all three LEDs should be observed. Yet, since the arrow head is pointing to the LED in the middle, some users thought that the arrow signified that only the middle LED should be observed, and quite rightly so, because the other LEDs at the top of the same diagram use a system of surrounding the group of LEDs with a rectangle, without the arrow head.

To sum up the comments from all test users, many agreed that a suitably balanced combination of textual and diagrammatic mode, enriched with a short animation and some easily understandable symbols here and there would possibly work the best with the material of this study. With some of the small screen steps, it was considered hard for a still picture or a diagram alone to convey all the information that the users expected, especially where does a step start and how one can conclude that a step is completed successfully. There was a frequent need to use the “Next” and “Back” buttons to check previous and upcoming steps, to understand the context of the workflow. However, it was not unanimous whether all steps should have a picture, diagram or animation of some kind. Some test users suggested that they would like to see a picture in all steps, while others were happy that, for example, there was no picture for connecting a familiar HDMI cable.

5.3 Implications on future research

As discussed in this thesis, producing effective small screen instructions encompasses such a broad field of topics to consider that this thesis was only able to scratch the surface. There were many interesting avenues opening up during the conversion production, and especially during the discussions with the test users, but I had to leave many of them out to limit the scope of this thesis. This section summarises these areas for further research.

The technical production environment is a clearly distinguishable area that should be studied further. This should include considering single-source content creation and publishing with, for example, DITA XML (or similar system), and extend to user interface design for small screens. As discussed in Section 5.1.1, it would benefit to publish the same source content in full form to larger medias, and in small screen format to devices having smaller screens, thus eliminating the need to duplicate the content for the screen size specialisation. The target would be an easy-to-manage content management system that publishes the content for the user in an optimal way.

As discussed in Section 2.4.3, Siltanen and Heinonen (2020, p. 107) call for more multimodal user interface development and further research. As indicated also by this thesis, users have personal preferences how they would like to work with instructions and equipment, so that the signified meanings carry over in the most effective way. The task at hand and the space where they are working in can limit the multimodal meaning-making affordances that the used medium has to offer. As stated earlier in this thesis, the 3D working situation can have an influence on how the message in the instructions carries over to the technician. For example, noise, lighting conditions, other people passing-by, even differences in the actual equipment as compared to what is depicted in the instructions can have an influence and should be addressed in further research.

Even though there are many designs in use already, further studies should find an *industrial, automated* method for taking more advantage from the emerging potential in augmented reality, where the role of the user can change into an interactive participant in a 3D spatiotemporal environment. For example, the augmented reality device could match its real time camera information to the image in the instruction: a real KONE MediaPlayer in a real lobby could be identified based on the KONE MediaPlayer image in the instructions, and the AR device would enable locking augmented reality arrows and graphics to the camera feed. The technology and such solutions exists, but they are often case specific and remade per each solution. In such an industrial authoring solution, adding audio for reading the instructions aloud, giving sound or haptic feedback for warnings, should be studied. All in all, the use of more semiotic modes than attempted in my design experiment, combined with the technical advances of today’s mediums, should be researched and tested in practise.

Providing context for small screen instructions should be studied more. As noted in the user tests, the users feel more confident when they know where they are coming from and where they are going to in regards of the overall continuum of the work they are conducting. Studying the context has a link to the user interface design on a small screen. Also not discussed in this thesis is the problem of findability, in the real context where technicians are working. Even if a company has thousands of instructions in its database, the instructions do not help the technicians unless they can find the right piece of instruction, on the exact time when they need it. This moment often occurs on site, next to the equipment, in a challenging work situation, with a possibly poor network connection.

Designing and understanding symbols should be studied in the context of technical communication, for example, what kind of symbol is understood best, do you teach the symbols beforehand, how quickly users learn the symbols, and how well they remember them. For example, Scurati et al. (2018, p. 77) state that the symbols with highest homogeneity are not necessarily the ones with highest guessability. To understand the challenges with symbols in today's global business setting, cultural differences in understanding symbols should be studied further. Here, theories of multimodal communication provide useful tools (see Section 2.4).

Lastly, continuing the emphasis on today's global setting, the social semiotics and cultural multimodality should be studied more regarding multimodal meaning-making in (small screen) instructions, and whether Industry 4.0 and digital globalisation have had any effect on the social and cultural differences in the context of technical communication.

6 Conclusions

This thesis examined how to convert professional installation and maintenance instructions for digital small screens, with a special focus on the users of the instructions from the semiotic, multimodal meaning-making point of view. During a hands-on design experiment, two theories, visual manual (Gattullo et al., 2019) and minimalism heuristics (van der Meij and Carroll, 1998), were followed to convert two originally A4 PDF-format installation tasks and one maintenance task to the small screen of RealWear HMT-1 smart glasses. However, from the content creation point of view, the small screen could have been any, for example, a mobile device, some other smart glasses, or even a smart watch. This thesis considered that for business reasons, such production of small screen instructions should also fit the traditional PDF production of a company, which somewhat limited exploring the multimodal affordances of the smart glasses. However, there is still a need for traditional PDFs or paper documents in the operations of a global company, so they cannot be neglected. The small screen conversions were then analysed with the theories of multimodality, focusing on the changes in the multimodal meaning-making, with a special focus on the users of the small screen conversions. With the users in mind, the analysis of the meaning-making changes was cross-checked with a qualitative user test, where test users performed the converted tasks on real equipment on a simulated working environment.

The results of the thesis were that the multimodal meaning-making changed, especially with the visual manual, but not extensively, since not all the modes available for meaning-making in the test medium were utilised due to not fitting into the scope of this thesis. With the visual manual theory, creating and understanding the symbols was found to make the biggest impact on both the authoring and the multimodal meaning-making. The more visual elements and modes were combined into the same step, the more work was required from the users. As an example, the grammatical reading of sign clusters was seen as a challenge. With minimalism heuristics, the results did not differ significantly from traditional instruction authoring and reading, provided that the existing KONE style guide is somewhat based on minimalism anyway. The biggest change was leaving out some obvious steps, for example, the reinstallation steps of a removed component, when in theory, the reinstallation steps could be easily deducted from the removal steps presented already. The opinions of the test users differed regarding whether they want to see all steps carefully documented, or whether they prefer implementing the minimalism heuristics so that obvious information can be left out. This thesis showed that this had an effect on the meaning-making, since some users got confused if the "rhythm" of the step-flow changed in the middle of the task, for example, by combining two steps into one, or leaving out obvious steps by saying "re-install in reverse order".

To conclude, this thesis noted that the potential area of research was too vast for one study to cover. Multiple areas for further research were identified in Section 5.3, with a high recommendation to pursue researching and developing a professional content management system with *fitting authoring guidelines* to accommodate both traditional paper-format instruction and small screen instruction production into the same system, thus eliminating the need for multiple work on the same content. The research should accommodate exploring not only the new possibilities offered by new digital devices, but also their use in the multimodal situation, where technicians conduct their work. Lastly, designing and understanding symbols in a global setting should be studied more in the context of technical communication.

And, with the help of theories of multimodal communication, whatever and how gets implemented on the small screens of Industry 4.0 instructions, we have systematic tools for analysing how the meanings get across to the users of the instructions.

Bibliography

Albers, M. 2002. Information design for the small-screen interface: An overview of web design issues for personal digital assistants. *Technical Communication*, vol. 49, no. 1, pp. 45–60.

Bateman, J.A. 2018a. The Communicative Effectiveness of Education Videos: Towards an Empirically-Motivated Multimodal Account. *Multimodal Technologies and Interaction*, vol. 2, no. 3, pp. 59–86.

Bateman, J.A. 2018b. Peircean Semiotics and Multimodality: Towards a New Synthesis. *Multimodal Communication*, vol. 7, no. 1.

Bateman, J.A. 2017. Triangulating transmediality: A multimodal semiotic framework relating media, modes and genres. *Discourse, Context & Media*, vol. 20, pp. 160–174.

Bateman, J.A. 2011. The Decomposability of Semiotic Modes. In: B.A. Smith and K. O'Halloran eds. 2011. *Multimodal Studies: Exploring Issues and Domains.* Routledge. pp. 17–38.

Bateman, J.A. 2008. *Multimodality and genre: A foundation for the systematic analysis of multimodal documents.* Basingstoke: Palgrave Macmillan.

Bateman, J., Wildfeuer, J. and Hiippala, T. 2017. *Multimodality. Foundations, research and analysis. A problem-oriented introduction.* Berlin/Boston: De Gruyter.

Bellamy, L., Carey, M. and Schlotfeldt, J. 2012. *DITA Best Practices: A Roadmap for Writing, Editing, and Architecting in DITA*. Upper Saddle River, N.J: IBM Press.

Bezemer, J. and Kress, G. 2016. *Multimodality, learning and communication: a social semiotic frame.* Milton Park: Routledge.

Carroll, J.M. 1990. *The Nurnberg funnel: Designing minimalist instruction for practical computer skill.* Cambridge, MA: MIT Press.

Churchill, D. and Hedberg, J. 2008. Learning object design considerations for small-screen handheld devices. *Computers & Education*, vol. 50, no. 3, pp. 881–893.

Deppermann, A. 2013. Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics*, vol. 46, no. 1, pp. 1 -7.

Engelke, T. 2013. Content first - A concept for industrial augmented reality maintenance applications using mobile devices. In: ACM, *MMSys '15: Proceedings of the 6th ACM Multimedia Systems Conference.* Portland, OR, USA, 18–20 March 2015. New York, NY, United States: Association for Computing Machinery. pp. 105–111.

Erkoyuncu, J.A., del Amo, I.F., Mura, M.D., Roy, R., Dini, G. 2017. Improving efficiency of industrial maintenance with context aware adaptive authoring in augmented reality. *CIRP Annals,* vol. 66, issue 1, pp. 465–468.

Eschen, H., Kötter, T., Rodeck, R., Harnisch, M. and Schüppstuhl, T. 2018. Augmented and Virtual Reality for Inspection and Maintenance Processes in the Aviation Industry. *Procedia Manufacturing*, vol. 19, pp. 156–163.

Gattullo, M., Scurati, G.W., Fiorentino, M., Uva, A.E., Ferrise, F. and Bordegoni, M. 2019. Towards augmented reality manuals for industry 4.0: A methodology. *Robotics and Computer Integrated Manufacturing*, vol. 56, pp. 276–286.

Gattullo, M., Uva, A.E., Fiorentino, M., Scurati, G.W. and Ferrise, F. 2017. From Paper Manual to AR Manual: Do We Still Need Text?. *Procedia Manufacturing*, vol. 11, pp. 1303–1310.

Gallon, R. and McDonald, A. 2016. Context Sensing and Information 4.0. *TC World*, November 2016, pp. 14–18.

Gimeno, J., Morillo, P., Orduña, J.M., Fernández, M. 2013. A new AR authoring tool using depth maps for industrial procedures. *Computers in Industry*, vol. 64, issue 9, pp. 1263–1271.

Hallgren, C. 1992. The Nurnberg Funnel: A minimal collection. *Journal of Computer Documentation*, no. 16 (1), pp. 11–17.

Hiippala, T. 2016. Aspects of Multimodality in Higher Education Monographs. In: A. Archer and E.O. Breuer eds. 2016. *Studies in Writing: Vol. 33, Multimodality in Higher Education*. Leiden: Brill. pp. 53–78.

Hiippala, T. 2015. *The structure of multimodal documents: An empirical approach.* New York: Routledge.

Hiippala, T. and Bateman, J.A. 2020. Introducing the diagrammatic mode. arXiv preprint: https://arxiv.org/abs/2001.11224

Horn, R. 1992. Commentary on the Nurnberg Funnel. *Journal of Computer Documentation.* no. 16(1), pp. 3–10.

Information Mapping 2020. *Information Mapping*. [online] Available at: <https://www.informationmapping.com> [Accessed 1 March 2020].

Jappy, T. 2013. *Introduction to Peircean visual semiotics.* London: Bloomsbury.

Jayaprakash, S. 2008. *Technical writing.* Mumbai: Himalaya Pub. House.

Kaasinen, E., Aromaa, S., Väätänen, A., Mäkelä, V., Hakulinen, J., Hella, J., Keskinen, T. and Turunen, M. 2018. Mobile Service Technician 4.0: Knowledge-Sharing Solutions for Industrial Field Maintenance. *Interaction Design and Architecture(s) Journal*, no. 38, pp. 6–27.

Kagermann, H., Wahlster, W., Helbig, J. and Hellinger, A. 2013. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group.* Frankfurt/Main: Forschungsunion.

Kittredge, R.I. 2005. Sublanguages and Controlled Languages. In: R. Mitkov ed. 2005. *The Oxford Handbook of Computational Linguistics*. Oxford handbooks online. pp. 430–448.

Knopfle, C., Weidenhausen, J., Chauvigne, L., Stock, I. 2005. Template based authoring for AR based service scenarios.In: IEEE, *IEEE Proceedings. VR*

2005. Virtual Reality, 2005. Bonn, Germany, 12–16 March 2005. New York, NY, United States: IEEE. pp. 237–240.

Kuhn, T. 2014. A Survey and Classification of Controlled Natural Languages. *Computational Linguistics*, vol. 40, no.1, pp.121–170.

Lodor, E. 1999. The proven and potential promises of minimalism for technical communication. *ACM SIGDOC Asterisk Journal of Computer Documentation*, vol. 23, no. 2, pp. 48–56.

Matthiessen, C.M. 2007. The multimodal page: a systemic functional exploration. In: T. D. Royce and W. Bowcher eds. *New Directions in the Analysis of Multimodal Discourse.* New York: Lawrence Erlbaum Associates. pp. 1–62.

Nickerson, R. 1991. A Minimalist approach to the "paradox of sense making”. *Educational Researcher.* vol. 20, no. 9. pp. 24–26.

Orwell, G. 1987. *Nineteen eighty-four.* London: Secker & Warburg.

Radkowski, R. 2015. Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. *International Journal of Human-Computer Interaction.* vol. 31, no. 5, pp. 337–349.

RealWear 2019. *RealWear*. [online] Available at: <https://realwear.com/> [Accessed 1 December 2019].

Redish, J. 1998. Minimalism in Technical Communication: Some Issues to Consider. In: J.M. Carroll ed. 1998. *Minimalism Beyond the Nurnberg Funnel.* Cambridge, Massachusetts: The MIT Press. pp. 219–246.

Rosenbaum, S. 1998. Follow-up Training in Minimalism: How are Technical Communicators Using Minimalism? In: J.M. Carroll ed. 1998. *Minimalism Beyond the Nurnberg Funnel.* Cambridge, Massachusetts: The MIT Press. pp. 119–148.

Schriver, K.A. 1997. *Dynamics in document design: Creating text for readers.* New York: Wiley.

Scurati, G.W., Gattullo, M., Fiorentino, M., Ferrise, F., Bordegoni, M. and Uva, A.E. 2018. Converting maintenance actions into standard symbols for Augmented Reality applications in Industry 4.0. *Computers in Industry*, vol. 98, pp. 68–79.

Siltanen, S. and Heinonen, H. 2020. Scalable and responsive information for industrial maintenance work: developing XR support on smart glasses for maintenance technicians. In: ACM, *Proceedings of Academic Mindtrek '20.* Tampere, Finland, 29–30 January 2020. New York, NY, USA: ACM. pp. 100–109.

Society for Technical Communication 2019. *Defining Technical Communication*. [online] Available at: <https://www.stc.org/about-stc/defining-technical-communication/> [Accessed 6 October 2019].

Stock, I., Weber, M., Steinmeier, E. 2005. Metadata based authoring for technical documentation. In: ACM, *SIGDOC05: ACM 23rd Annual International Conference on Documentation*. Coventry, United Kingdom, September 2005. New York, NY, USA: ACM. pp. 60–67.

Tham, J., Duin, A.H., Gee, L., Ernst, N., Abdelqader, B. and McGrath, M. 2018. Understanding Virtual Reality: Presence, Embodiment, and Professional Practice. *IEEE Transactions on Professional Communication,* vol. 61, no. 2, pp. 178–195.

Tripp, S. 1990. Review of the Nurnberg Funnel. *Educational Research and Development*, vol. 38, no. 3, pp. 87–90.

Waller, R. 2012: Graphic Literacies for a Digital Age: The Survival of Layout. *The Information Society*, vol.28, issue 4, pp. 236–252.

van der Meij, H. and Carroll, J.M. 1998. Principles and Heuristics for Designing Minimalist Instruction. In: J.M. Carroll ed. 1998. *Minimalism Beyond the Nurnberg Funnel.* Cambridge, Massachusetts: The MIT Press. pp. 19–54.

Virtaluoto, J., Suojanen, T. and Isohella, S. 2021. Minimalism Heuristics Revisited: Developing a Practical Review Tool. Preprint for *Technical Communication*, February 2021.

Wildfeuer J., Pflaeging J., Bateman J., Seizov O. and Tseng C. 2019. *Multimodality: Disciplinary Thoughts and the Challenge of Diversity.* De Gruyter.

Wobbrock, J.O., Aung, H.H., Rothrock, B., Myers, B.A. 2005. Maximizing the guessability of symbolic input. In: ACM, *Extended Abstracts on Human Factors in Computing Systems, CHI EA '05.* Portland, Oregon, USA, 2–7 April 2005. New York, NY, USA: ACM. pp. 1869–1872.