# AR Authoring: How to Reduce Errors from the Start?

Camille Truong-Allié[1][a], Martin Herbeth[2] and Alexis Paljic[1][b]
[1]*Centre de Robotique, Mines Paris - PSL, Paris, France*
[2]*Spectral TMS, Évry-Courcouronnes, France*

Keywords: Augmented Reality, Task Assistance, AR Authoring Interface, Knowledge Transmission.

Abstract: Augmented Reality (AR) can be used to efficiently guide users in procedures by overlaying virtual content onto the real world. To facilitate the use of AR for creating procedures, multiple AR authoring tools have been introduced. However, they often assume that authors digitize the procedure perfectly well the first time; this is yet hardly the case. We focus on how AR authoring tools can support authors during the procedure formalization. We introduce three authoring methods. The first one is a video-based method, where a video recording is done before procedure digitization, to improve procedure recency, the second one an in-situ method, where the digitization is made in the procedure environment, to improve context, and the last one is the baseline method, where AR authors digitize from memory. We assess the quality of the procedures resulting from these authoring methods with two simple yet underexplored metrics: the number of errors and the number of versions until the final procedure. We collected feedbacks from AR authors in a field study to validate their significance. We found that participants' performance was better with the video-based method, followed by the in-situ and then the baseline methods. The field study showed the advantages of the different methods depending on the use case and validated the importance of measuring digitization error.



## 1 INTRODUCTION

Augmented Reality (AR), with its ability to overlay virtual content to real-world elements, has already proven its usability and efficiency in instructing users about procedures they have to perform (Fite-Georgel, 2011). For example, Head Mounted Displays (HMDs) enable users to keep their hands free. With this, users can perform a task while following AR instructions superimposed to the physical environment such as animated 3D models or videos.

A procedure is typically a series of steps that need to be completed to achieve a goal. In an industrial context, a procedure is often a description of these steps which is made available to workers as digital documents or on paper. A well-written AR procedure should ensure workers safety and efficiency.

To help virtual content creators (AR authors) to design these procedures, multiple AR authoring tools have been proposed. Figures 1 and 2 are examples of one authoring tool. They enable an AR author with no programming skills to create a software which provides AR instructions (Kearsley, 1982). It could be possible, for example, to type text and arrange pictures or 3D models in 3D space without code. Almost all of the works about AR authoring tools consider that the AR authors know perfectly well the procedure, how to explain it best and digitize it perfectly right the first time - this is often not the case on a daily basis. Rather, they more likely loop between authoring the content and reviewing it from the operator view (Scholtz and Maher, 2014; Gerbec et al., 2017). . It is for instance common to forget a step, or to realize while following the AR-assisted procedure that another formulation would have been clearer.

In this work, we are interested in the authoring tool conditions which can help best AR authors to formalize procedures. We propose three methods related to the moment AR authors plan the procedure digitization: a method such that they perform the procedure before digitizing it (video-based method, where AR authors first capture themselves executing the procedure); a method such that the procedure digitization is at the procedure location (in-situ method), and a method where AR authors are left alone when digitizing the procedure (baseline method, where AR authors digitize the procedure off-site, by memory).

We start by reviewing the different authoring tools

[a] https://orcid.org/0000-0002-0973-7004
[b] https://orcid.org/0000-0002-3314-951X



Truong-Allié, C., Herbeth, M. and Paljic, A.
AR Authoring: How to Reduce Errors from the Start?.
DOI: 10.5220/0012303200003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 408-418
ISBN: 978-989-758-679-8; ISSN: 2184-4321

that we found in the literature, and how they have been evaluated. Then, we introduce the three authoring methods above-mentioned and compare them in a user study. Finally, we draw conclusions, limitations and future work.

# 2 RELATED WORK

## 2.1 Difficulties in Authoring Tools

We are interested in authoring tools specifically designed for procedural tasks, enabling the digitization of AR procedures as step-by-step instructions. Existing AR authoring tools focus on easing AR content creation, AR content placement, or procedure organization.

### 2.1.1 AR Content Creation

AR content creation can be time-consuming and require specific skills, for example to create complex 3D models (Gattullo et al., 2019). To help AR authors to create AR content without the need of specific skills, three main strategies exist. 1. Some authoring tools propose a database of AR content (typically, 3D models) in which AR authors can pick the appropriate content (Knopfle et al., 2005; Blattgerste et al., 2019). 2. Some authoring tools automatically create the AR content from the capture of AR author gesture or their environment with computer vision (Chidambaram et al., 2021; Pham et al., 2021) or of a product to assemble in the case of assembly (Zogopoulos et al., 2022). 3. Finally, some AR authoring tools rely on simple AR content which does not require high skills to be created: video, picture, text, and simple 3D models like arrows (Lavric et al., 2021; Blattgerste et al., 2019).

### 2.1.2 AR Content Placement

Another difficulty when creating AR content is that it should be integrated into the real-world. Manipulating the AR content and place it accurately is not an easy task. Multiple works propose to help AR authors in the AR content placement. It is possible to associate physical markers to virtual representations that serve as references for the placement of AR content. These representations represent either directly the virtual elements to display in AR (Zauner et al., 2003), either the real environment, in which the virtual content can be placed on a desktop application (Gimeno et al., 2013). It is also possible to place the virtual content with 2D interfaces on devices that AR authors may be familiar with: a computer, with positioning of the virtual content in a 2D or 3D graphical interface (Zauner et al., 2003; Haringer and Regenbrecht, 2002; Knopfle et al., 2005; Bégout et al., 2020), a mobile phone, in a joystick controller fashion (Blattgerste et al., 2019). Finally, some authoring tools automatically position the virtual content, based on author positioning in the real-world (Chidambaram et al., 2021) or automatic scene analysis (Pham et al., 2021; Erkoyuncu et al., 2017).

### 2.1.3 Procedure Organization

Creating and placing the AR content is not enough to make AR procedures. AR authors need to organize its structure so that the whole procedure is coherent. To help AR authors organizing the procedure, some tools propose an automatic segmentation of the procedure into steps, and one tool proposes a bi-directional system which enable operators to correct errors AR authors could have made in the procedure.

In the context of assembly, it is possible to automatically detect the assembly steps as they all have a similar structure that consists of the addition of parts to assemble. Step detection can be done from a digital twin of the final product to assemble (Zogopoulos et al., 2022) or computer vision (Bhattacharya and Winer, 2019; Funk et al., 2018). For tasks which are more complex than assembly, the Ajalon tool enables automatic step detection at the cost of a higher authoring tool complexity (Pham et al., 2021). This tool helps AR authors to organize their procedure in a finite-state-machine from which an adaptive software is derived. The resulting software automatically detects the step the operator is performing, gives the corresponding instructions and indicates if the step is wrongly performed. Finally, ACARS (Authorable Context-aware Augmented Reality System) helps AR authors in the creation of AR procedure, not with an automatic step segmentation, but with a bi-directional system that enables operators to point errors that have been made in the procedure (Zhu et al., 2013). This tool consists of two parts: an offline authoring tool for AR authors, that enable them to create a context-aware software from static rules, and an in-situ authoring tool for operators in which they follow the AR instructions, interact with it and update it. The static rules enable to elect the right content depending on the detected input context (i.e. choose level of detail depending on expertise level).

### 2.1.4 Conclusion

The related work shows that, while multiple works propose simple tools for AR content creation and placement, fewer works are proposed when it comes to organizing it in a coherent structure. They are based on automatic step segmentation, which is today challenging and limited to simple tasks like assembly, or specific applications and cannot be generalized. To the best of our knowledge, ACARS is the only tool that considers AR authors fallibility and enable operators to correct their errors. This feature is major yet underexplored in the literature. In this work, we focus on an earlier phase of the authoring process. We aim to prevent these errors from occurring in the first place, rather than simply correcting them afterward.

We are interested in how AR authoring tools can be designed to best help AR authors in the procedure organization. We consider the moment when they formalize their procedure, and question how to organize the authoring tool around this moment. To do so, we propose tools which can improve memory recall. Recall is facilitated by practice, recency and context (what is present in the person's focus of attention) (Budiu, 2014). We propose two authoring methods. One is a video-based authoring method with which the AR author first captures a first-person video of the procedure before they formalize it - it is designed to improve recall by recency. The other is an in-situ authoring method, where the AR author formalizes the procedure at the location of the procedure, enabling them to observe the procedure environment and, if desired, even perform the procedure - it is designed to improve context.

## 2.2 Evaluation Methods

Our work focuses on how good the digitized procedure is. We therefore need to understand what a good procedure means. To answer this, we take interest in how non AR procedures and AR authoring tools have been evaluated.

### 2.2.1 Non AR Procedure Evaluation

Traditional paper procedures can be evaluated in terms of risk of human error (Kirwan, 1997; Gertman et al., 1992; Gertman et al., 2005) or complexity (Park and Jung, 2007).

Human Reliability Analysis methods have been proposed to evaluate human error probabilities given a procedure. They all rely on an expert analysis of the procedure and/or its environment, leading to subjective measures and possible inconsistencies (Jang and Park, 2022) and a time-consuming analysis for a large set of procedures. TACOM, which stands for TAsk COMplexity, is a measure which gives a score of task complexity from quantifiable measures (Park and Jung, 2007). The measures require knowledge about the task context and a time-consuming analysis. For example, for each step of the task, logic complexity and information complexity should be evaluated. To palliate the time-consuming analysis required by Human Reliability Analysis methods and TACOM scores, machine learning and natural language processing-based algorithms have been proposed to evaluate procedures complexity based on their structures (Sasangohar et al., 2018; McDonald et al., 2023).

### 2.2.2 Authoring Tools Evaluation

Works on AR authoring tools propose different methods for their evaluation. The evaluations mainly consider AR author experience, the quality of the resulting AR procedure by considering the operator experience when following it, and, if the authoring tool has an automated part, its error rate. AR author experience is usually measured in terms of procedure creation time, cognitive load and usability of the tool. Operator experience is evaluated in terms of performance (task completion time and error rate), cognitive load and usability of the resulting procedure. Table 1 summarizes the different AR authoring tools evaluation methods used by the previous work. It only includes the authoring tools that have been evaluated.

### 2.2.3 Conclusion

There is no straightforward method to evaluate the quality of a procedure organization, and the existing methods focus on whether a procedure is clear rather than correct.

The evaluation of traditional non AR procedures aims to clarify an existing procedure which is already correct. It either requires complex analysis skills, either machine learning analysis that can be difficult to implement. The evaluation of AR authoring tools indirectly assesses the resulting procedure quality with operators experience when following it. No metric directly related to the procedure quality is used. Finally, at the exception of ACARS, all the existing AR authoring tools start from a finalized procedure, where the only concern is to be digitized with AR. They do not consider the process in which the procedure is created and improved before reaching its final state. Zhu et al., with ACARS, propose a method to correct authoring errors, but they did not evaluate how this method improves the final procedure.

In this work, we propose two simple metrics to evaluate the quality of a digitized procedure organization: the number of authoring errors made by AR authors until they are satisfied with the final procedure,

Table 1: Existing AR authoring tools and their evaluation: With AR author, operator or tool performances. TCT stands for Task Completion Time, RT for Reading Time, QMC for Quantity of Manually Created Content, ER for Error Rate, EP for Error in Positioning, CQ for Custom Questionnaire, SUS for System Usability Scale, UEQ for Usability Experience Questionnaire, NX for NASA-TLX, R for Recall, P for Precision, IoU for Intersection over Union (for object detection), TT for Training/Testing Time (for machine learning model).

| Eval. / Paper | AR author | | Operator | | Tool |
|---|---|---|---|---|---|
| | Objective | Subjective | Objective | Subjective | Objective |
| (Blattgerste et al., 2019) | | SUS, UEQ | TCT, ER | SUS, NX | |
| (Chidambaram et al., 2021) | TCT | CQ, SUS, NX | TCT, ER | CQ, SUS, NX | |
| (Pham et al., 2021) | TCT, QMC | CQ | | | P, R, IoU |
| (Erkoyuncu et al., 2017) | TCT | | TCT | | |
| (Lavric et al., 2021) | | | TCT, RT, ER | CQ, SUS | |
| (Gimeno et al., 2013) | TCT, EP | CQ | TCT | | |
| (Bégout et al., 2020) | TCT | | | | |
| (Bhattacharya and Winer, 2019) | | | | | P, R, TT |
| (Funk et al., 2018) | TCT | NX | TCT, ER | NX | |
| (Zhu et al., 2013) | | | | CQ | |

and the number of versions they make to achieve this final procedure. We used these metrics to compare three different authoring methods, and gather AR authors feedback in a field study to validate the importance of these two metrics.

## 3 AUTHORING METHODS

The authoring methods described in this section are inspired from the idea that recall can be improved by recency and context. They are all based on the same two applications: a desktop and an AR applications.

### 3.1 Baseline

The baseline authoring method consists in alternating between the desktop application, in which AR authors set and describe all the elements to constitute the procedure, and the AR application on HMD, where AR authors can visualize the elements and place them in the real-world (Spectral TMS, ).

The desktop application enables to first write the whole procedure, step by step. Each step can have a title, a textual description, pictures, videos and 3D models attached to them. We chose these virtual content types as they are the most used and preferred ones in industries (Gattullo et al., 2020).

After writing the procedure, the AR authors can use the AR application to place the virtual content within the physical environment. If errors are noticed during this step, they have to go back to the desktop application to correct them in a new version. Then, they use the AR application again to place any new virtual content and verify that no error is left, etc.

### 3.2 Video-Based Authoring

The video-based authoring method is such that AR authors capture themselves performing the task from a first-view perspective with a HMD. Then, they use the desktop authoring tool described in the baseline and the video they just made as a support to write a first version of the procedure. The authoring steps are then the same as the ones proposed in the baseline method.

This method can be extended into realistic authoring tools, like Taqtile (Taqtile ), or Ajalon (Pham et al., 2021), where the expert video is the main medium to create the whole AR procedure. For example, the video can be used for content creation and placement with automatic object detection and procedure organization can be made even easier with automatic step segmentation.

### 3.3 In-situ Authoring

With the in-situ authoring method, AR authors are at the procedure location when writing the procedure on the desktop application. The steps are then the same than in the baseline but at the procedure location. This way, authors can look at the procedure elements and even perform the procedure to improve their recall.

This method can be extended into more complex authoring tools that rely on the in-situ location of the AR author, for example, by making possible to create virtual content on the spot, like pictures or videos (Lavric et al., 2021; Blattgerste et al., 2019).

# 4 EXPERIMENT

## 4.1 Objective

Each of the methods we proposed aims to represent a diverse range of authoring tools; we design the methods and the corresponding authoring tool to be as generic as possible. What we needed was for the three methods to be comparable in a way that we are able to measure the effects of their characteristics alone: for the video-based method, the effect of adding video capture before digitization, and for the in-situ method, the effect of having the AR author physically present on-site.

The objective of this experiment is to capture the effect of such characteristics on participants procedure creation time, number of errors, number of versions until the final one, and cognitive load.

## 4.2 Variables

We measured procedure creation time, video capture included for the video-based authoring method. We used NASA-RTLX (Byers, 1989) to assess participants' cognitive load at the end of the experiment. The score was calculated from linear scales from 1 to 100. We measured the number of errors and the number of versions before the final version of the procedure. A version corresponds to the state of the procedure after a loop between the edition (mainly on desktop application) and review of the procedure (on AR application). The number of errors between two consecutive versions is the number of changes made by the participant between the two versions. The total number of errors is the sum of all the changes between consecutive versions. This means that each participant evaluates themselves whether there was an error in their procedure by correcting it. We deemed this to be the most effective method for error measurement since the definition of an error is subjective and therefore tricky to evaluate. For instance, one person might perceive a step as too simple to mention, while another might view this same step omission as an error. By using the number of changes between two versions, we measured the number of self-detected errors, that is, what participants actually consider an error.

## 4.3 Participants

30 participants took part in the experiment, 7 women and 23 men. They came from diverse firms and from a computer science research lab. They were 29 on average - std 3. When asked about their familiarity with AR and the task to perform (make a coffee with a capsules machine) from 0 to 4, they ranked on average their familiarity with AR to 1.4 (std 1.7) and with the coffee machine to 2.6 (std 1.5).

Some participants where not familiar with the coffee machine; they were consequently not given the baseline condition.

## 4.4 Experimental Setup

**Procedure to Digitize.** Participants' task was to create a procedure to explain how to make a coffee with a capsule machine with one of the three authoring methods (between-subject study). This task was chosen because it is relatively common so that many people can be considered as experts, but still complex enough for participants to make mistakes. The coffee machine had its water container empty (that needs to be filled) and its old capsules container overfilled (that needs to be emptied) for each participant: those were the two main causes of errors.

**Desktop Application.** Because the experiment does not focus on virtual content creation, we provided participants with dummy templates for the virtual content they could use if needed: pictures, videos and 3D models. Those were dummy media with an icon representing what the actual media should have been: a camera for the picture, a video camera for the video and cube in a reference frame for the 3D model (see what the templates look like in Figure 1). In addition to these virtual content types, participants could also indicate physical elements location by placing 3D spheres, as described by Figure 2. The virtual content and physical elements locations can be created by clicking on the "Add an equipment" and "Adding media" buttons.

**AR Application.** On the AR application, all the virtual content can be placed in the physical environment; Figure 2 shows two examples of virtual content for two different steps. The title and textual description are displayed on a dashboard that participants can pin wherever they like. The AR part of the authoring tool is not evaluated in the experiment and is the same for all the authoring methods.

## 4.5 Procedure

Participants were first explained how the authoring tool worked by creating a very simple procedure (draw a smiley on a whiteboard). They created the steps on the desktop application, and then, they watched a tutorial video explaining how to use the

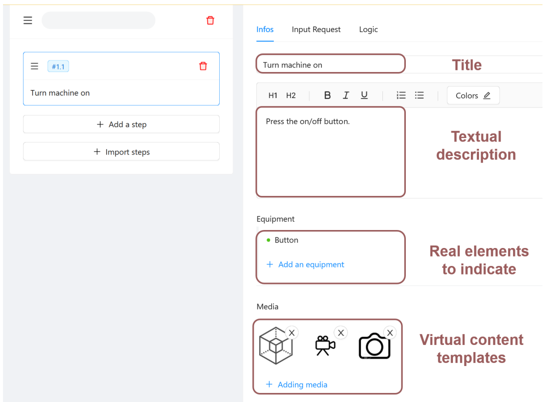


Figure 1: The authoring tool desktop application, where it is possible to create procedure steps, describe them with a title, a text, real-world elements locations, pictures, videos and 3D models.

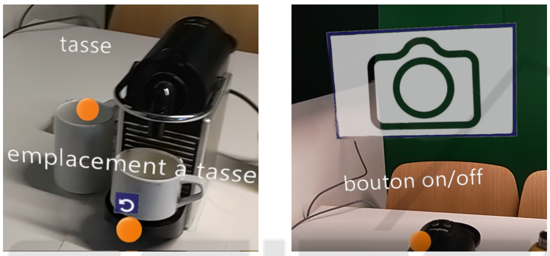


Figure 2: Two examples of the virtual content that is displayed on the AR application. On the left are indications of real-world elements with orange spheres. On the right is a dummy picture that the participant placed as if it were the actual picture they wanted to use.

AR application. After this, they realized the actual experiment with one of the three authoring methods. They created their procedure on the desktop application, placed the virtual content and reviewed it on the AR application, corrected it on desktop in a new version if needed, etc. They were asked to stop when satisfied. Finally they answered the NASA-RTLX questionnaire and were free to leave comments about the experiment.

## 4.6 Data Analysis

When the assumptions for the one-way ANOVA were met (mainly normality and homoscedasticity), we used it to analyze the data, otherwise, nonparametric Kruskal-Wallis test was used. The procedure creation time and NASA-RTLX data are normal (Shapiro's test p-value are respectively 0.7 and 0.9), continuous and homoscedastic (Levene's test p-value are 0.4 and 0.1). We could perform one-way ANOVA for these variables. The number of versions and number of errors are both categorical data; we used Kruskal-Wallis.

## 4.7 Results

### 4.7.1 Effect of Authoring Methods

Participants performed best with the video-based authoring method for all the metrics, as shown in Figure 3, although no statistically significant results were found (see Table 2).

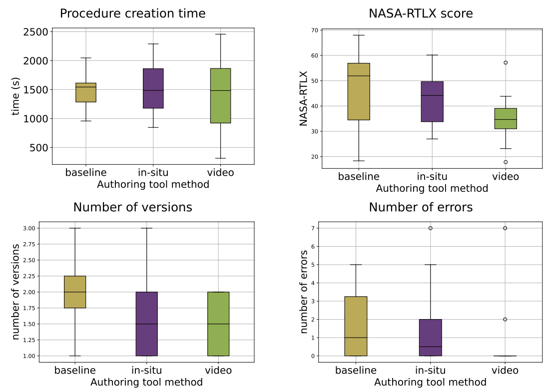


Figure 3: Boxplots of participants results between the different authoring methods. Procedure creation time is the time taken by participants to reach the final version of the procedure, video recording included for the video-based method. NASA-RTLX assesses cognitive load (Byers, 1989). The number of versions is the number of versions participants made before reaching the final procedure. The number of errors is the number of errors they self-detected within their procedure before reaching the final procedure.

Table 2: P-Values of the Kruskal-Wallis tests for the effect of authoring method on procedure creation time, NASA-RTLX, number of versions and number of errors.

| Measure | P-value |
|---|---|
| Procedure creation time | 0.74 |
| NASA-RTLX score | 0.27 |
| Number of versions | 0.67 |
| Number of errors | 0.56 |

### 4.7.2 Effect of Expertise

In Section 4.7.1, we did not consider the effect of participants' expertise on their performances. Indeed, we found in a prior analysis that adding participants' expertise in the data modeling did not significantly improve it. This result was obtained with the ANOVA function from the R language, by comparing a model with participants' expertise and one without.

Yet, when separating participants between novices (coffee expertise under 3) and experts, different trends were observed between the two groups for the procedure creation time. This is the only variable in which two clearly different trends can be observed. In general, novices created the procedure faster with the in-situ method than with the video-based method, but experts created the procedure slower with the in-situ method than with the video-based and the baseline methods, as illustrated by Figure 4.

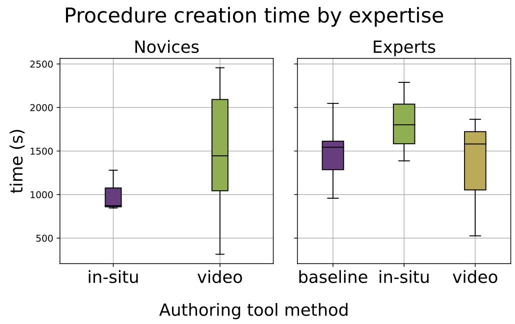


Figure 4: Boxplots of procedure creation times by authoring methods and participants' expertise. Baseline was not proposed to novices as they were not able to write a procedure without a video or being in-situ.

## 4.8 Conclusion

Participants' performances are best (although not significantly) for all the metrics with the video-based authoring method, followed by the in-situ authoring method and finally the baseline. Novices and experts have similar trends except for time data, where novices are faster with the in-situ method than with the video-based method.

## 4.9 Discussion and Limitations

Participants had better results with the video-based method. Even with the video capture time included in the procedure creation time, participants were faster with the video-based than with the in-situ method. This means that enforcing that AR authors, even experts, perform the procedure before digitizing it makes them earn time. This is probably due to the video forcing them to perform the task, while with the in-situ authoring method, they were free to do as they pleased. We indeed noticed that only a few participants actually performed the procedure with the in-situ method; most of them mimicked it or mentally reviewed the steps to perform by looking at the environment. This characterizes well what a real-use would be, as we told participants to do as they wanted.

On the contrary, novices took more time with the video-based than with the in-situ method. This can be explained by the video recording being a learning moment for novices rather than a reminder. When recording the video, novices just discovered the task, this probably was not enough to make a mental version of the procedure to perform. Novices might have started to formalize it only while using the desktop application. And, there, contrary to the in-situ condition, they did not have the procedure environment as a reminder. More work would be required to verify this hypothesis.

The data was slightly biased because no novice was assigned the baseline. This should have had only a limited impact on our results as we shown that experts and novices shared similar trends most of the time. Additionally, what interests us most is the difference between video-based and in-situ, and novices were equally distributed between these conditions.

## 4.10 Future Works

An interesting future work would be to improve the video-based and in-situ methods that we proposed. In this direction, it could be interesting to better understand the role of recall within the different methods. In our work, we did not specifically assess how recency and context affected recall. Investigating these factors could provide insights into why the video-based method outperformed the in-situ one and guide the development of AR authoring tools that more effectively exploit the benefits of recency and/or context. This could be done for instance by varying the degrees of context and recency provided to participants when digitizing a specific procedure.

It could also be interesting to analyze the evolution of participants' knowledge during procedure digitization. It would enable to identify the moments when participants formalize the procedure, and compare how fast participants acquire or recall knowledge in the different authoring conditions. Knowledge could be subjectively assessed by grades of participants' confidence in their capacity to digitize the procedure from memory, or objectively, with questionnaires about the procedures.

In this experiment, we measured procedure quality in terms of errors that are self-detected by AR authors and of number of versions they made before being satisfied with the procedure. These two metrics have the advantage of being simple to measure and to straightforwardly represent a limitation of the AR authoring tool, but they do not measure how well a procedure is organized. It would be interesting to measure procedure complexity using the metrics and tools detailed in Section 2.2.1.

Finally, in this work, we focused on recall, but other cognitive processes that help AR authors formalize a procedure could have been considered, e.g. mental imagery, logical reasoning or analytical thinking. They would have led to other possible designs. As an example, mental imagery involves creating mental representations of concepts, objects, or processes (Bronkhorst et al., 2020), and it has been shown that providing thematic content improves mental imagery and therefore problem solving skills (Clement and Falmagne, 1986). Consequently, it could be interesting to design authoring tools that propose thematic content (e.g. elements useful for the procedure) to improve mental imagery and logical reasoning.

# 5 FIELD STUDY AT INDUSTRIAL SITES

## 5.1 Objectives

After the experiment described in Section 4, we performed a field study where several AR authors from 7 different industrial sites and 3 firms were interviewed. The field study had two objectives. The first objective was related to the authoring methods: we wanted to make sure that they are viable and represent concrete use of AR authoring tools. The second objective was related to the metrics we used to compare the authoring methods (number of errors and versions): we wanted to verify their actual interest in the assessment of the quality of an authoring tool.

## 5.2 Setup

During the interviews, we asked AR authors how they digitize their procedure. We asked them the context and use case for which they digitize a procedure, the existing materials that they use to digitize the procedure, and their method to digitize the procedure. We then question them about the video-based and in-situ methods: if they are interesting for them and in which conditions. Finally, we asked if they needed multiple versions before reaching the final procedure. AR authors digitize their procedure using the two applications described in Section 3.1. To complete digitization, they need: to create AR content (text, picture, video or 3D model); to organize their procedure within steps; to place the AR content. During the interviews, we focused on AR content creation and procedure organization.

## 5.3 Results

### 5.3.1 Authoring Process

**Digitization Context.** AR authors digitize procedures in different contexts and for different reasons. Two of them are procedure creators: when new products or machines are created, they design and digitize the AR procedure to assemble or use them. Three of them digitize existing procedures with the help of already existing materials (but possibly out-of-date). Finally, two of them digitize existing procedures but without any prior material. The summary of how AR authors use the authoring tool is given by Table 3.

**Expert Video Prior Digitization.** Three out of seven AR authors create an expert video prior to procedure digitization. They use it for two reasons: making sure that they digitize an updated procedure (the expert video is the most recent performance of the procedure), and creating AR content for the procedure: chunks of the expert video for each step or annotated screenshots.

**AR Content Creation in-situ.** Five out of seven AR authors go in-situ prior and during procedure digitization to create visual asset; primarily pictures and videos.

### 5.3.2 Authoring Methods

All of the AR authors use the video-based, the in-situ, or both methods. Those using the expert video use it to limit digitization errors and to create AR content. Those going in-situ do it to create AR content.

Multiple AR authors mentioned that they would like to use the expert video to create all the virtual content because it would prevent them from needing to go in-situ. Yet, they often raised concerns about the expert video as a good medium for content creation. Indeed, they said that it required skills and time to create a high quality expert video which can be used for media creation. Making the expert video at the right angle is difficult. Sometimes, what needs to be captured is small and would require zooming. The use of a HMD to make a video requires practice: often, the camera jitters, the AR authors need to rotate the head in an unnatural position to capture the right elements, which can also be hidden from the AR author point-of-view.

The creation of AR content in-situ can be time-consuming, often because AR authors need several trials before being satisfied with the content. For example, they can realize while editing the expert video on desktop that the expert forgot to wear the security equipment and that they need to do the video all over again.

### 5.3.3 Numbers of Errors and Versions Metrics

All of the AR authors mentioned high number of errors and versions before reaching the final procedure. The number of versions is even higher when the AR

Table 3: Summary of AR authors' digitization methods gathered during a field study. Each industrial site uses AR for a specific use case. They can digitize the procedure from existing materials, record an expert's video of the procedure before digitizing it, and create AR content (pictures, videos) at the location of the procedure (in-situ).

| Site | Use cases | Existing materials | Expert's video prior digitization | Media creation in-situ |
|---|---|---|---|---|
| 1 | Procedures for new products | × | × | ✓ |
| 2 | Procedures for new machines | × | × | ✓ |
| 3 | Quality control procedures | × | × | ✓ |
| 4 | Assembly procedures | × | ✓ | × |
| 5 | Assembly, quality control procedures | ✓ | ✓ | × |
| 7 | Format shift, diagnoses procedures | ✓ | × | ✓ |
| 6 | Digitization of spreadsheets procedures | ✓ | ✓ | ✓ |

authors create a procedure which does not exist yet: they have to write drafts that they heavily modify at each new version. The creation of media also induces multiple versions: AR authors first write the procedure, then go at its location to create media that they then have to edit. They regularly do this several times, for example to recapture an unsatisfactory media.

## 5.4 Conclusion and Discussion

### 5.4.1 Authoring Methods

This field study validated the video as a tool for AR authors to do less authoring mistakes because it forces them to digitize with the the up-to-date procedure in mind. It additionally highlighted the difficulty of creating high quality AR content, and suggested two possibilities to ease the creation of content: either using the expert video as a base for other assets, either by privileging in-situ authoring tools. The former has the drawback of relying on video creation, which requires skills, and the latter has the drawbacks of not effectively reducing authoring errors, and of potentially disturbing production lines for an undetermined period for industrial applications.

### 5.4.2 Number of Errors and Versions Metrics

This field study highlighted that the number of versions is both due to digitization errors and to AR content creation. While in this case, the number of versions represents yet another difficulty of authoring tool (media creation), it does not enable to capture procedure quality alone. The number of errors is a more precise metric to evaluate the procedure quality.

This field study showed that both metrics are relevant and and concern AR authors on a daily basis. We argue that they should be used in the design of authoring tools in two aspects. The first aspect is AR authoring tool evaluation during its design: these metrics enable to assess the quality of the procedures resulting from the authoring tool. The second aspect relates to the design of the authoring tool itself. This study showed that it is very unlikely that AR authors digitize their procedure in a single version. Not only can they make mistakes, but they can also want to improve the quality of the AR content, or the procedure can change through time and AR authors need to update it. If they create a new procedure, they work iteratively until they are satisfied with it. We argue that AR authoring tools that enable a smooth versioning and editing would improve AR authors experience. For example, works could focus on making the edition of AR content simple, even for assets requiring high technical skills like video or 3D models.

# 6 CONCLUSION

In this work, we studied how AR authoring conditions can improve the quality of the resulting procedure. To evaluate this quality, we measured the number of self-detected authoring errors and number of versions until AR authors are satisfied with their procedure. We compared three authoring conditions, a video-based method, an in-situ method, and a baseline method. We found that the video-based method outperformed the other methods in terms of the two metrics above-mentioned as well as in terms of procedure creation time and cognitive load. In addition, in a field study, we gathered AR authors' feedback about how they digitize AR procedures and what they think of the authoring methods above-mentioned. Depending on the use case for which AR is needed, they favoured a different authoring method. They all produced a high number of versions before being satisfied with their procedure, highlighting the relevance of self-detected authoring errors and number of versions metrics.


# ACKNOWLEDGEMENTS

The authors wish to thank all the participants of the experiment: thank you for your time, we hope you enjoyed it.

## REFERENCES


Bégout, P., Duval, T., Kubicki, S., Charbonnier, B., and Bricard, E. (2020). Waat: A workstation ar authoring tool for industry 4.0. In *Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part II 7*, pages 304–320. Springer.

Bhattacharya, B. and Winer, E. H. (2019). Augmented reality via expert demonstration authoring (areda). *Computers in Industry*, 105:61–79.

Blattgerste, J., Renner, P., and Pfeiffer, T. (2019). Authorable augmented reality instructions for assistance and training in work environments. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*, pages 1–11.

Bronkhorst, H., Roorda, G., Suhre, C., and Goedhart, M. (2020). Logical reasoning in formal and everyday reasoning tasks. *International Journal of Science and Mathematics Education*, 18:1673–1694.

Budiu, R. (2014). Memory recognition and recall in user interfaces. *Nielsen Norman Group*, 1.

Byers, J. C. (1989). Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary? *Advances in Industrial Erfonomics and Safety l: Taylor and Francis.*

Chidambaram, S., Huang, H., He, F., Qian, X., Villanueva, A. M., Redick, T. S., Stuerzlinger, W., and Ramani, K. (2021). Processar: An augmented reality-based tool to create in-situ procedural 2d/3d ar instructions. In *Designing Interactive Systems Conference 2021*, DIS '21, page 234–249, New York, NY, USA. Association for Computing Machinery.

Clement, C. A. and Falmagne, R. J. (1986). Logical reasoning, world knowledge, and mental imagery: Interconnections in cognitive processes. *Memory & Cognition*, 14:299–307.

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

Fite-Georgel, P. (2011). Is there a reality in industrial augmented reality? In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 201–210.

Funk, M., Lischke, L., Mayer, S., Shirazi, A. S., and Schmidt, A. (2018). Teach me how! interactive assembly instructions using demonstration and in-situ projection. *Assistive Augmentation*, pages 49–73.

Gattullo, M., Evangelista, A., Uva, A. E., Fiorentino, M., and Gabbard, J. L. (2020). What, how, and why are visual assets used in industrial augmented reality? a systematic review and classification in maintenance, assembly, and training (from 1997 to 2019). *IEEE transactions on visualization and computer graphics*, 28(2):1443–1456.

Gattullo, M., Scurati, G. W., Evangelista, A., Ferrise, F., Fiorentino, M., and Uva, A. E. (2019). Informing the use of visual assets in industrial augmented reality. In *International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering*, pages 106–117. Springer.

Gerbec, M., Balfe, N., Leva, M. C., Prast, S., and Demichela, M. (2017). Design of procedures for rare, new or complex processes: Part 1 – an iterative risk-based approach and case study. *Safety Science*, 100:195–202. SAFETY: Methods and applications for Total Safety Management.

Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C., et al. (2005). The spar-h human reliability analysis method. *US Nuclear Regulatory Commission*, 230(4):35.

Gertman, D. I., Blackman, H. S., Haney, L. N., Seidler, K. S., and Hahn, H. A. (1992). Intent: a method for estimating human error probabilities for decision-based errors. *Reliability Engineering & System Safety*, 35(2):127–136.

Gimeno, J., Morillo, P., Orduña, J., and Fernández, M. (2013). A new ar authoring tool using depth maps for industrial procedures. *Computers in Industry*, 64(9):1263–1271. Special Issue: 3D Imaging in Industry.

Haringer, M. and Regenbrecht, H. (2002). A pragmatic approach to augmented reality authoring. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pages 237–245.

Jang, I. and Park, J. (2022). Determining the complexity level of proceduralized tasks in a digitalized main control room using the tacom measure. *Nuclear Engineering and Technology*, 54(11):4170–4180.

Kearsley, G. (1982). Authoring systems in computer based education. *Commun. ACM*, 25(7):429–437.

Kirwan, B. (1997). The development of a nuclear chemical plant human reliability management approach: Hrms and jhedi. *Reliability Engineering & System Safety*, 56(2):107–133.

Knopfle, C., Weidenhausen, J., Chauvigne, L., and Stock, I. (2005). Template based authoring for ar based service scenarios. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pages 237–240.

Lavric, T., Bricard, E., Preda, M., and Zaharia, T. (2021). Exploring low-cost visual assets for conveying assembly instructions in ar. In *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6.

McDonald, A. D., Ade, N., and Peres, S. C. (2023). Predicting procedure step performance from operator and text features: A critical first step toward machine learning-driven procedure design. *Human Factors*, 65(5):701–717. PMID: 32988239.

Park, J. and Jung, W. (2007). A study on the development of a task complexity measure for emergency operating procedures of nuclear power plants. *Reliability Engineering & System Safety*, 92(8):1102–1116.

Pham, T. A., Wang, J., Iyengar, R., Xiao, Y., Pillai, P., Klatzky, R., and Satyanarayanan, M. (2021). Ajalon: Simplifying the authoring of wearable cognitive assistants. *Software: Practice and Experience*, 51(8):1773–1797.

Sasangohar, F., Peres, S. C., Williams, J. P., Smith, A., and Mannan, M. S. (2018). Investigating written procedures in process safety: Qualitative data analysis of interviews from high risk facilities. *Process Safety and Environmental Protection*, 113:30–39.

Scholtz, C. R. and Maher, S. T. (2014). Tips for the creation and application of effective operating procedures. *Process Safety Progress*, 33(4):350–354.

Spectral TMS. The technician's augmented reality assistant. https://www.spectraltms.com/en/. Accessed: 2021-12-17.

Taqtile Knowledge when and where you need it. https://taqtile.com/. Accessed: 2023-09-14.

Zauner, J., Haller, M., Brandl, A., and Hartman, W. (2003). Authoring of a mixed reality assembly instructor for hierarchical structures. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pages 237–246.

Zhu, J., Ong, S. K., and Nee, A. Y. (2013). An authorable context-aware augmented reality system to assist the maintenance technicians. *The International Journal of Advanced Manufacturing Technology*, 66:1699–1714.

Zogopoulos, V., Geurts, E., Gors, D., and Kauffmann, S. (2022). Authoring tool for automatic generation of augmented reality instruction sequence for manual operations. *Procedia CIRP*, 106:84–89. 9th CIRP Conference on Assembly Technology and Systems.