

PRE-PRINT

Real-time Coding Method and Tool for Artefact-centric Interaction Analysis in Co-design Situations Assisted by Augmented Reality

<https://link.springer.com/article/10.1007/s12008-020-00683-8>

C. Masclet¹, M. Poulin¹, J.F. Boujut¹, N. Becattini²

1- Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000 Grenoble, France

2- Politecnico di Milano - Department of Mechanical Engineering, Milano

* Corresponding author. Tel.: +33682175390. E-mail address: jean-francois.boujut@grenoble-inp.fr

Abstract

This paper presents and discusses a fast and efficient method to study artefact-centric interactions in co-design sessions. The method is particularly useful for exploring the introduction of augmented reality (AR) environments since AR application representations combine both digital content and physical content, transforming the way users interact with the design object. Although protocol analysis is extensively used in cognitive studies of design, it is a time-consuming and cumbersome method and hence unsuitable for extensive analysis in industrial environments. Our real-time coding method makes it possible to perform “on-the-fly” coding of physical interactions in co-design sessions. Focusing on quantifying interaction occurrences, our results are consistent with those obtained with post-session coding. Internal validity was assessed using relevant statistical tests. Based on the data collected in co-design sessions we show how aggregate results, especially timelines and interaction densities, can be displayed rapidly at the end of each session. This research paves the way for a more general implementation of real-time coding of collaborative work sessions in industrial situations.

Keywords

Co-design; collaborative design; real-time coding; artefact-centric interaction; gesture analysis; augmented reality; mixed artefact

Co-creativity involving designers and end users can substantially contribute to design company competitiveness [1]. During the design process, co-creative sessions provide insights into a product's use allowing potential end users to express ideas and give direct feedback to the designers. In today's information and communication society an increasing number of users are keen to take an active part in the understanding, modification or even creation of the products they use [2]. As part of this new approach, designers and end users collaborate by proposing, testing and assessing alternative product ideas before finally validating the chosen design. The aim of co-creative sessions is to take advantage of what Steiner calls "process gain" [3]. This is the phenomenon according to which the overall design outcome is improved through the early integration of end users. The assumption made by the promoters of this kind of session is that the complementary vision of end users and designers, backed up by the designers' special skills, contributes to the success of product development activities. Tools have been developed and tested [4] that show a significant impact of the interactivity functions on the outcome of the design and the perception of the users [5]. These approaches are considered to be part of the interactive design paradigm that develop today in the companies.

However, end users are not familiar with the representations and tools used by designers in their daily work. While designers comfortably handle digital product representations in order to define shapes and dimensions (CAD models or rendering tools), or simulate product behaviour (FEM analysis, kinematic behaviour, etc.), end users often prefer physical, realistic objects that allow them to assess subjective parameters such as ease-of-use, aesthetics and what is more generally called the user experience. The involvement of multiple senses appears to be an advantage for the evaluation of alternative solutions [6]. This often leads to the coexistence of two types of representation: digital and physical. Hence, collaboration is hindered by the fact that each stakeholder prefers to work with different types of design representation. As a consequence, design loops are extended due to the introduction of physical prototypes that are labour-intensive and expensive to produce.

Today, the participation of end users in the early design phases is considered critical by many design agencies. This means that processes, tools and practices must evolve to take into account the end users' need to see and feel objects. Introducing augmented reality technology allowing real-time modifications of a realistic product representation could facilitate interaction between designers and end users, thus enhancing end users' participation in co-creative sessions [4]. The SPARK project (<http://spark-project.net/>), funded by the European Union's Horizon 2020 Research and Innovation program, aims to do precisely this. The project provides a responsive Information and Communication Technology (ICT) platform based on Spatial Augmented Reality (SAR) technology. This ICT platform supports the creation, visualisation, assessment and modification of design concepts from the earliest

stages of a design process. SAR was introduced by [7] as the projection of a virtual digital object on to a physical shape. It is also called “projection mapping” because projectors are used to display additional, digital content on the real-world object (figure 1).

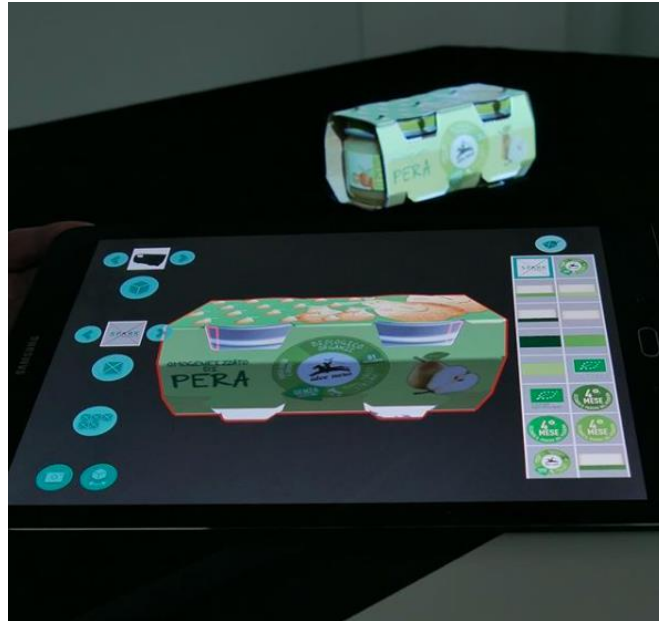


Figure 1: View of the Spatial Augmented Reality prototype and the tablet-based user interface

SAR visualisation creates a mixed object [8]: it has the appearance and tangibility of a physical object but its surface is modified by the display of digital content. Additionally, a tracking functionality allows participants to handle the mixed object as easily as they would a real physical object. Hence end users can experience a prototype that is as close as possible to the real thing. This tangible interaction allows them to be more involved in the design activity; they are able to express their expectations and feeling based on an artefact that helps them to illustrate their statements [9]. From the designers' perspective, this technology enables rapid changes to the digital content. The design can be swiftly and smoothly modified in real time according to the end users' expectations. The technology does not impair the flow of interactions between the stakeholders affording the designers better control over the pace of the co-design sessions.

In order to study the influence of these mixed artefacts on the interactions between end users and designers during co-design sessions and to determine if this influence is beneficial to the overall session outcomes, we need a robust and reliable method to capture and analyse the large amount of data produced in these complex industrial situations (figure 2).



Figure 2: Typical co-design situation (<https://stimulo.com/>)

The method and tool proposed here aim to help researchers to:

- Assess the distribution of interactions between end users and designers and quantify their respective participation;
- Evaluate the type of artefact (e.g. physical, digital or mixed) used during these interactions;
- Display the density of interactions so that the most interesting episodes can be identified in view of more in-depth cognitive studies.

The overall approach aims to contribute to existing cognitive approaches by providing results based on directly observable physical interactions.

In what follows, the method for collecting and quantifying artefact-centric interactions in various co-design situations shall be presented and validated. The preliminary results show the typical outcomes of the method and illustrate the variety of analyses that can be performed.

State of the art

1.1. Design protocol analysis

Design protocol analysis is the application of the protocol analysis method [10] for analysing design activity and understanding the designers' thought processes. [11] coordinated a research effort based on an experiment known today as the "Delft protocol", in reference to workshops held at Delft University of Technology where the approach was first presented. This was an important attempt to articulate diverse points of view, based on different theoretical frameworks, using a single protocol.

A design protocol is a set of rules for a design task and procedure enabling the needs of a research question to be met. The protocol lays out the organisation of the task or procedure and can be set up in a lab context or captured in an industrial setting. Its analysis requires recording designers' speech and behaviour together with any material produced during the design session. According to [12], video capture is *a fundamental technique for constituting the corpora of data for analysis, as well as an important means of rendering research results*. This method first enables researchers to capture data verbalisation, sketches, drawings, physical mock-ups, etc., in other words all the audio-visual events occurring during a design session. It then allows them to repeatedly listen to verbal exchanges and watch and review designers' actions as part of an in-depth post-session analysis.

Many of the early studies based on design protocol analysis focused on verbal reports of designers thinking aloud [13-16]. Design studies using design protocol analysis primarily aim to capture and understand relevant design cognitive processes [17] or elementary activities. Examples of activities identified from design protocol analysis studies include: analysing problems, identifying criteria, proposing solutions or syntheses, assessing or analysing solutions and making decisions [18-21]. As design protocol analysis became more widespread, researchers began to take into account the collaborative dimension of design. A significant range of studies concentrated on groups of designers working in small teams [22-26]. [27] proposed a taxonomy to differentiate between these two types of design protocol: individual, "think-aloud" protocols versus "group discussion" protocols. In a review of papers published between 1975 and 2006, [28] highlight the significant increase in use of design protocol analysis during this period. They point to the co-existence of these two types of study while underlining the predominance of the think-aloud type. Since the study presented here focuses on co-creative design sessions, the group discussion protocol method is the most appropriate approach.

Identifying designers' cognitive models and intentions requires a substantial amount of work in terms of speech modelling and analysis. However, for co-design sessions, where communication plays an important role, interactions between the participants are a major observable source of data for researchers before they turn their attention to more complex cognitive activities. Interactions may be supported by artefacts, gestures or both. Therefore, studying artefact centric interactions could provide an objective measure of the intensity of the collaboration without requiring the analysis of cognitive activities, if we make the assumption that artefact centric gesture interactions are the visible part of more complex cognitive interactions among the design participants.

1.2. Artefact-centric Interaction Analysis

When researchers study collaborative activities, such as design meetings, the objects shared and produced as part of the design interactions are seen to be important for the whole process [29].

Likewise, gestures are considered as a complementary means of expressing ideas and conveying meaning during collaborative design sessions. Stakeholders move their hands and other parts of their body for communicative purposes [30]. [31] underlines how, as conventional design tasks are being performed, gestures often accompany speech. Indeed, repeated gestures may even be shared and integrated into the team's "conceptual lexicon" becoming part of the team's kinaesthetic thinking, as highlighted by [32]. Other investigations in the field of communication and gestural explanations refer to McNeill's categorisation of gestures [33] as iconic, deictic and metaphoric. Based on McNeill's work [33], [34] coded the gestures of learning session participants as they explained what they had understood about the circulatory system. The researchers counted the number of gestures associated with each category, noted when the participants began to move their hands or returned them to their initial position and how the participants emphasised points by increasing the magnitude of their movements.

During co-design sessions both physical and digital artefacts play an important role in gestural interactions. According to [35] and [36], designers spend most of their time during their design activity creating, handling, discussing, interpreting, evaluating and transforming texts, graphs, calculations, drawings and digital or physical mock-ups. These authors propose and discuss the concept of the "Intermediary Object" (IO). This term lends the objects and artefacts handled by designers an appropriate status for analysing and understanding their design activity. IO's have three main features: they are representations of the object to be designed (a means of making the object visible); they are involved in a complex translation process (from knowledge to its crystallisation into the object); and, finally, they act as mediators between the various actors (hence supporting interactions). These objects are considered "open" [37] when they can be modified by the designers during design meetings or "closed" when they convey non-ambiguous content. SAR technology provides tangible objects with digital content that is open to modification. Mixed objects can be considered as "open intermediary objects" since the digital content of these objects can be altered in real time. Capturing the three main types of interaction occurring around these artefacts provides us with objective information about each actor's participation and the type of artefacts used. For verbal interactions this involves noting who speaks or initiates the interaction. For graphical interactions it entails noting who draws and on what. Finally, for gestural interactions the researcher focuses on who handles what, who observes what and who points at what.

Hence, studying artefact-centric interactions appears to be a relevant scientific method for studying the role of artefacts in collaborative design sessions and, in the more specific context of this research, the impact of new augmented reality technology on the co-creation process.

1.3. Artefact-centric interaction framework

Our own coding scheme [38] focuses on gestures supporting artefact-centric interactions and considers three elements: the end user(s), designer(s) and artefact(s) involved in the interactions. We differentiate between two types of interaction: with artefacts (Tangible, Digital, Mixed and Ephemeral) and without artefacts (None) (Figure 3). This coding scheme was tested during design sessions with professional designers [38].

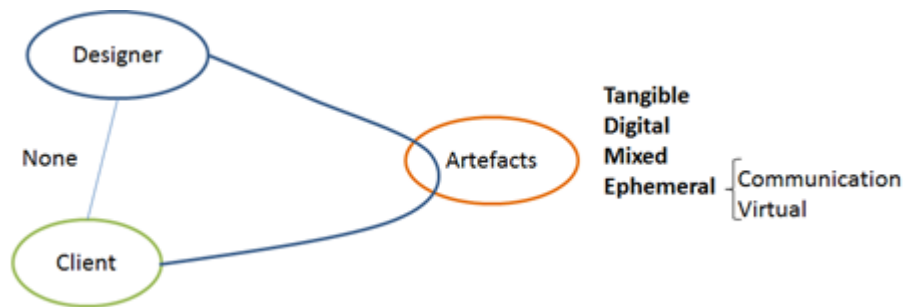


Figure 3. Artefact-centric interaction framework [38]

Tangible	Physical mock-ups, printed sheets outlining the alternative concepts, drawings, post-it notes
Digital	Representations displayed on a TV screen, laptop, tablet, smartphone
Mixed	Physical prototypes (physical mock-ups with a predefined shape, mostly 3D printed) on which digital elements are projected. These can be pictures, images, text or textures.
Ephemeral	Gestures in the air with no physical support

Table 1: Types of artefact considered in the analysis framework

We added the last “Ephemeral” category of “gestures in the air” to depict the numerous gestures performed by participants without any physical objects. Ephemeral artefacts include virtual artefacts and communicative gestures. The latter, such as “beat gestures” [32], complement speech and are not related to message content. A virtual artefact is defined as an imaginary object that is depicted or mimicked by a gesture in the air [38]. The person making the gesture can depict or mimic an object (shape, volume, surface), a usage (function in a specific context) or behaviour (deformation of an object, flashing light simulation, etc.).

Artefact-centric interaction analysis seems to offer considerable potential for exploring new co-design environments based on spatial augmented reality. The state of the art in this field reflects how important it is. Furthermore, given recent technological developments, a new observation approach based on gestural analysis is called for. However, in their review paper, [28] note a ratio of

analysis time over protocol length ranging from 10:1 to 1000:1, while [39] note that 25 hours of additional time per protocol is required for planning, transcription and analysis. In the next sections we put forward an on-the-fly coding method potentially capable of reducing the coder's burden while ensuring high quality results.

1.4. Research questions

Design protocol analysis is known to be labour-intensive [40] and time-consuming [10,28]. Additionally, the quality of the results depends on the time spent on analysis. By reducing the complexity of the phenomenon, we intend to capture preliminary data that would reduce the workload of the protocol analysis by allowing to focus on more intense episodes in terms of interactions. And, as we consider that the intensity of artefact-centric interactions are an observable of design collaboration, we can potentially conclude that we have a fast tool for eliminating non-significant episodes. Additionally, as mentioned in the introductory section, the new collaborative environments based on virtual and augmented reality require substantial interface validation and assessment work in industrial settings. On the other hand, collaborative environments are complex and therefore difficult to observe and reproduce using in vitro observations. This is why a large amount of data from real industrial settings needs to be captured to study the transformation of co-design within these new environments. In what follows, we shall address two research questions:

RQ 1: “Does a *real-time* coding method generate reliable data sets comparable to those obtained with a post-session coding method?”

RQ 2: “What kind of results can be extracted and automatically displayed from *on-the-fly* coding?”

Obtaining answers to these research questions would increase the possibility of doing more quantitative data analysis of complex design protocols. The method proposed in this paper is designed to act as a complementary approach to conventional design protocol analysis with the aim of reducing the overall coding effort.

In the next section we present the methodology used to generate quantitative data using on-the-fly coding. Internal validity is assessed by checking inter-coder reliability indexes and external validity through a statistical comparison of sets of quantitative data generated by on-the-fly coding and post-session coding. A critical analysis of our method is given and a suggested protocol to further develop this kind of methodology presented.

2. Methodology

2.1. Description of case studies and empirical material

2.1.1 Selection of case studies

The work presented here relied heavily on the observations of co-design sessions conducted within the SPARK European Project. The sessions considered involved design teams from two design agencies (Stimulo and Artefice¹). Five sessions were run under three different experimental conditions in Milan, Grenoble and Barcelona. The sessions covered the range of material configurations encountered in co-design sessions:

- The first configuration involved digital representations (3D models, slides, photos, etc.) and/or physical representations (post-it notes, plastic or cardboard prototypes). This configuration will be referred to as the *standard condition*;
- The second configuration involved the use of an Augmented Reality system featuring video see-through technology on tablets. This configuration is referred to as the *AR condition*;
- The third configuration involved the SPARK Spatial Augmented Reality platform. This configuration is referred to as the *SAR condition*.

The experimental room setting allowed us to control as many of the external parameters as possible: spatial distribution and profiles of participants, time, type of material available, technology, etc. The sessions brought together industrial case studies, professional designers and selected end users in order to be as close as possible to a real-life industrial situation.

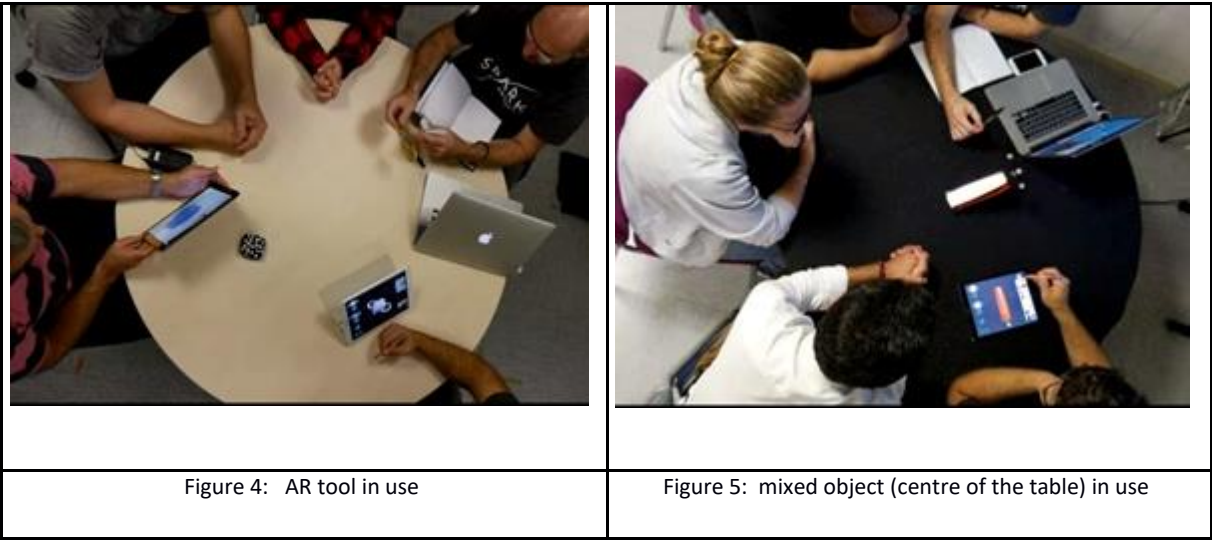
2.1.2. Case studies

The sessions used to validate the method lasted between 35 and 55 minutes and involved two designers from the partner design agencies as well as several potential end users. The sessions were organised and managed by the designers following their usual design process. No instructions were given by the research team. Each session began with the designers giving a short introduction to the product using PowerPoint slides. The co-creative activities began after this introduction. The aim was to develop the initial design proposals in light of the feedback and suggestions given by the participants. During the session the participants were allowed to use the tools available if they wished (AR tool, SAR tool or no specific tool).

¹These companies are part of the SPARK consortium. Artefice specialises in brand design while Stimulo designs innovative products.

The experimental setting for these configurations is illustrated in figures 4, 5 and 6 respectively. Figure 4 shows a tablet being handled by an end user (on the left). The tablet displays the AR scene of the object captured by the tablet's camera. The physical artefact is overlaid with a digital image representing the designed appearance of the product using video see-through technology. The tablet in the lower part of figure 5 displays the interface software used by the designers to modify the digital content projected on to the physical prototype (the illuminated artefact in the centre of the table).

Two coders standing in the experimental room performed coding using the on-the-fly tool (figures 7a and 7b) during the session.



Each session was recorded for the purpose of post-session coding. In all, four viewpoints were recorded to capture as much information as possible and each participant was equipped with a lapel microphone. Figure 6 is a screenshot showing the four video streams recorded during a standard session (tangible artefacts are available in the centre of the table).



Figure 6: Screenshot of the video processed from a standard session

2.2. The Observer tool and method

2.2.1 Description of the coding tool and scheme

The tool used by the coders was a software program providing a user interface that implemented the coding scheme (figures 7a and 7b). The coding scheme for the analysis comprised two categories: actor and artefact. Preliminary tests demonstrated that it was extremely difficult for a single coder to deal with both of these coding categories in real time. Due to the task complexity and the associated cognitive load on a single coder, we decided to divide the coding work between two coders. They produced separate primary data that were then merged in a later step. This merging step proved to be highly dependent on the way the coding had been performed. The more accurate the coding, the easier it was to merge the data (see section 2.2.2).

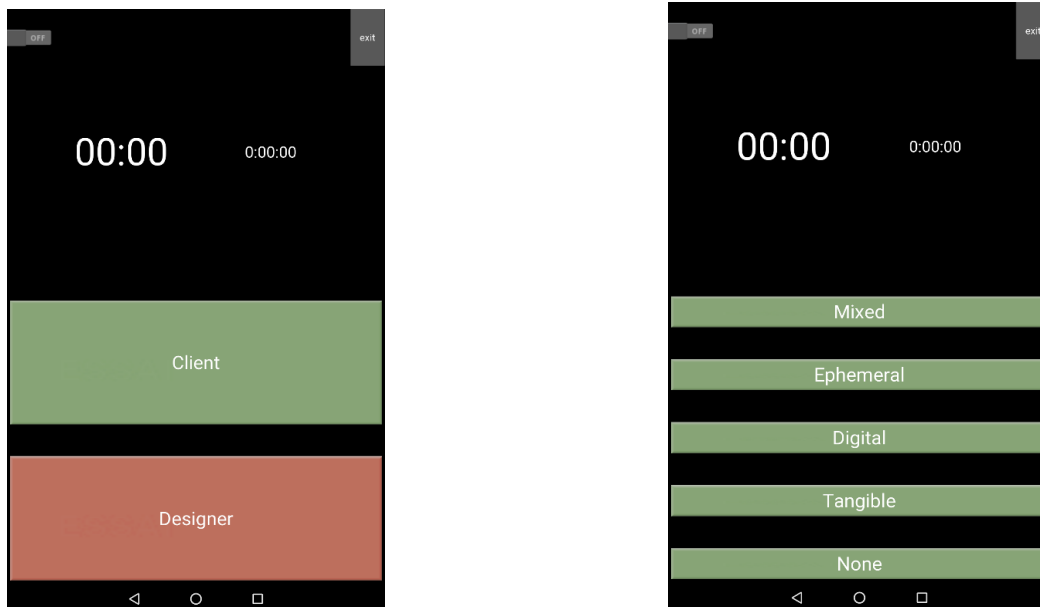
Since the project had a number of constraints (e.g. real work situations and non-intrusive observations), only two coders were allowed to be present in the session room for on-the-fly coding. This meant we had to share our ten coding items between two interfaces. These two complementary interfaces were developed with two main criteria in mind: the item had to be observable and the corresponding button easy to identify. The coders had to work in a real, complex environment and be able to capture all events, even those that were ambiguous or very brief and difficult to detect. To ensure reliable coding, regardless of the person performing it, we prepared a coding book describing visual clues and rules to facilitate the task. Section 2.3 describes how this part of the method was validated. One coder was put in charge of coding “the initiator of the interaction”, while the second

coder was responsible for capturing “which artefact supports the interaction” This decision was based on the idea that it would be easier for each human coder to focus on one aspect:

- the actor who is speaking (e.g. End user or Designer), using the voice as a clue;
- the main support for the interaction (e.g. Tangible; Digital; Mixed; Ephemeral; None), using the artefact as a clue.

This coding strategy was implemented via a graphic user interface (GUI) presenting two distinct layouts (figure 7 a and b) to the two coders. Once the *Observer* program was launched, the coders simultaneously started the recording by pressing the button in the upper left hand corner of the screen. They could then begin coding individually and independently. For ergonomic reasons, the buttons were larger than usual so that the coders could easily locate them using one or both hands. Here, the main concern was to enable the researchers to quickly alternate between observing the scene and looking at the interface without missing any events.

The primary data output from this coding step was the product of the two coders’ individual results. These results displayed the timestamps of the recorded events together with the type of event: in terms of 1) who was speaking and 2) the artefact supporting the interaction. Figure 8 below provides some examples of primary data from these two categories.



Figures 7a and 7b: *Observer software* interfaces for coding either actors (left) or interaction-supporting artefacts (right).

2.2.2 Data management and post-processing

The objective at this point in the process was to merge the primary data into a single file that faithfully represented the interactions occurring during the session, as illustrated in figure 8.

In order to match the two coding streams we followed two sub-steps:

- Sub-step “A” was automated by means of a software program that transformed the separate coding results provided by each coder into a first merged result draft. An excerpt of this first data arrangement is given in figure 8.
- Sub-step “B”, which aimed to obtain the final merging of results, required a manual operation to solve some of the remaining ambiguities following the automatic operation. The way this was achieved and the rationale for this strategy was described in the coding book. The book provided a set of simple rules to be followed based on common sense.

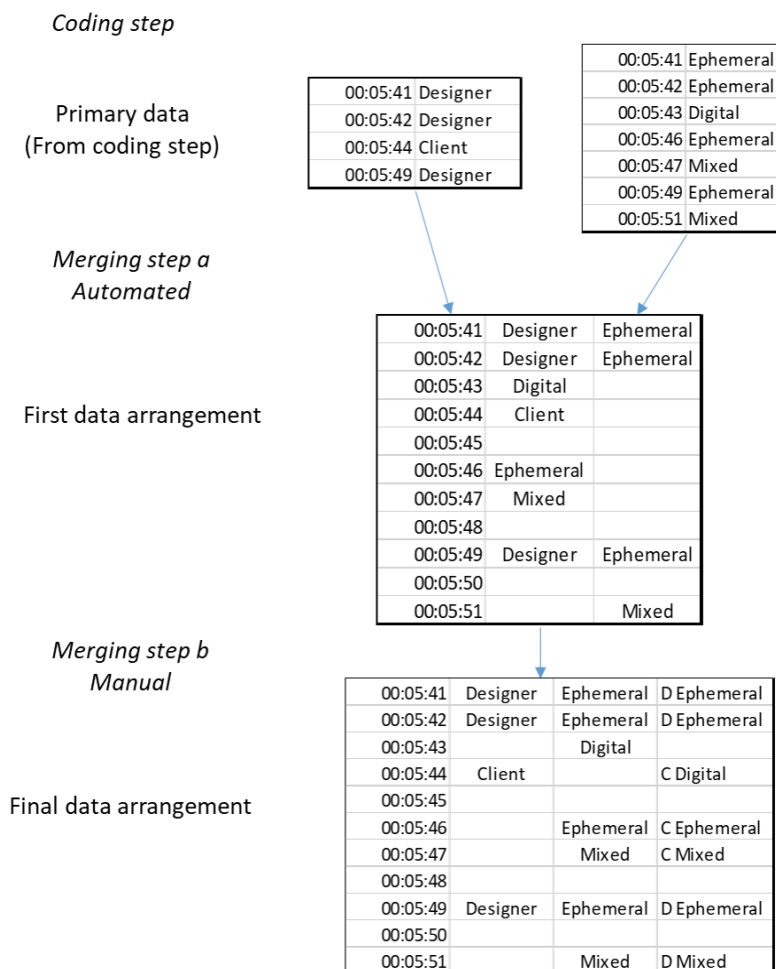


Figure 8: Procedure for merging the two coding files as part of the *on-the-fly* data coding method.

The coding book was developed to support the training of novice coders and provide expert coders with reminders before each new session.

2.3. Internal validation

Some preliminary tests allowed us to validate the usability of the *Observer* tool and the validity of the dual-coder strategy. In this section we shall demonstrate replicability of the results using an inter-rater reliability test.

2.3.1 Validation process and method

For the first test, two researchers were invited to code three 5-minute videos using the *Observer* tool. These videos were excerpts from co-design sessions with our industrial partners. The researchers were positioned in front of a computer with their tablet and were asked to code the sessions displayed on the screen in real time. Each coder coded the videos twice. The first time, they were asked to code the actor initiating the interaction (using the interface shown in figure 7a). The second time, they were asked to code the artefact used to support the interaction (using the interface shown in figure 7b).

This test allowed us to conclude that the *Observer* tool was easy to use for coding the initiating actor. It was also found to be suitable for coding the supporting artefact, although this required more cognitive effort owing to the five types of artefact.

When the two researchers were asked to code three samples from different sessions, the results showed a good level of convergence. However, a few disagreements were noted. These coding divergences were identified and discussed so that examples of coding difficulties could be compiled. To improve coding reliability, a coding book was developed based on the enhancements made to the initial coding modalities following review and discussion. Here are some examples of the coding rules laid out:

- Give priority to end users' interactions when designer and end user voices overlap;
- Even short verbal interactions, e.g. interjections such as 'OK' or 'Yeah', must be coded;
- Give priority to Ephemeral artefacts over other artefacts. For example, when an actor makes gestures before a Tangible, Digital or Mixed artefact, this should be coded as Ephemeral.

The coding book was checked and validated by the coding team.

2.3.2 Results of internal validation

With the coding book established, we proceeded to conduct a set of tests on a 5-minute sample video. The Cohen's Kappa associated with the coding of the new video excerpt performed by the same two coders is shown in the Table 1.

ON THE FLY	Standard condition		SAR condition		AR condition	
	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %	Cohen's Kappa	Agreement %
Actor	0.55	73	0.71	84	0.58	76
Artefact	0.62	74	0.59	71	0.30	49

Table 2: Cohen's Kappa for each session analysis

The Cohen's Kappa scale shows substantial agreement (green), moderate agreement (yellow) and fair agreement (red) between the two coders. The first excerpt from the AR session for artefact coding generates only a fair agreement level (0.30). This can be explained by the difficulty involved in coding artefacts. Artefact coding requires substantial training. However, we considered the level of convergence sufficient since the other two samples showed moderate to substantial agreement (SAR and Standard sessions). More tests would certainly have led to better results as the coders' experience was built up.

After having demonstrated the internal validity of the method using the *Observer* tool, we shall now compare the results of standard post-session coding with on-the-fly coding.

3. Results

3.1. Comparison of results from on-the-fly coding and post-session coding

This comparison is based on the two sets of final data obtained from the on-the-fly coding during live sessions at Stimulo and Artefice and the post-session coding. We propose two indicators:

- the percentage of actor interactions,
- the percentage of artefact involvement quoted for each category (e.g. digital, tangible, mixed, ephemeral, none).

The two sets of observations proved to converge well in terms of actor interactions when comparing on-the-fly coding with post-session coding (divergence from 0.4 % to 4.8 %) (Table 3), except for the Stimulo Standard condition (13.8 %, 18.4 %) and Artefice SAR condition for end users (8.5 %).

	Stimulo Session				Artefice Session			
	% of Designers' interactions		% of End users' interactions		% of Designers' interactions		% of End users' interaction	
	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session
Standard Condition	57.2	49.3	42.8	50.7	52.1	54.2	47.9	45.8
SAR Condition	66.9	67.2	33.1	32.8	66.1	63.2	33.9	36.8
AR Condition	63.9	63.6	36.1	36.4	44.6	46.3	55.4	52.7

Table 3: Percentage of actors' interactions during Stimulo and Artefice sessions

Regarding the percentages of gesture artefact-centric occurrences (see tables 4-6), the on-the-fly and post-session coding results are also very close. This strongly suggests that the on-the-fly coding approach is sufficiently relevant and accurate compared with the standard post-session coding approach.

	Stimulo Session				Artefice Session			
	% of Designers' interactions		% of End users' interactions		% of Designers' interactions		% of End users' interactions	
	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session
Digital	16.9	13.8	14.5	17.5	3.6	0.8	2.1	0.2
Tangible	15.1	12.7	12.3	16.4	42.5	41.8	33.2	29.4
Mixed	0	0	0	0	0	0	0	0
Ephemeral	18.0	14.5	9.6	7.5	3.6	6.2	6.8	10.4
None	7.2	8.4	6.4	9.3	5.0	5.4	3.2	5.8

Table 4: Artefact occurrences during the Stimulo and Artefice Standard session.

	Stimulo Session				Artefice Session			
	% of Designers' interactions		% of End users' interactions		% of Designers' interactions		% of End users' interactions	
	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session
Digital	1.7	2.5	0	0.2	6.5	2.0	3.6	5.0
Tangible	0	0	0.3	0	0.8	1.5	0.3	1.3
Mixed	38.6	36.5	20.3	19.2	38.6	43.1	23.6	22.4
Ephemeral	15.9	15.4	7.6	5.3	10.5	9.1	3.5	2.5
None	10.8	12.8	4.9	8.1	7.8	7.6	4.8	5.5

Table 5: Artefact occurrences during the Stimulo and Artefice SAR session.

	Stimulo Session				Artefice Session			
	% of Designers' interactions		% of End users' interactions		% of Designers' interactions		% of End users' interactions	
	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session	On-the-fly	Post-session
Digital	42.1	41.7	27.3	26.5	34.7	30.6	42.1	35.9
Tangible	3.6	2.2	1.4	1.7	0.4	0	0.3	0
Mixed	0	0	0	0	0	0	0	0
Ephemeral	12.3	13.9	4.3	4.7	8.0	11.4	5.8	11.8
None	6.0	5.8	3.1	3.5	5.0	5.3	3.8	5.0

Table 6: Artefact occurrences during the Stimulo and Artefice AR session

Tables 4, 5 and 6 all show that the percentages obtained with on-the-fly coding were very close to those obtained with post-session coding. However, the difference between on-the-fly and post-session results was sometimes as high as 8 to 9%. This divergence was found to be due mainly to the different distribution of coded events between the tangible artefacts (mixed, digital or tangible) and the non-tangible ones ("ephemeral" or "none"). A deeper analysis of the two coding results revealed that the

overall sum of coded interactions was equal for both coding approaches. This emphasises the need for coders to be properly trained to identify ephemeral gestures.

We shall now extend our analysis of the differences between these two coding methods by presenting the results of applying the Wilcoxon signed-rank test.

3.2. Wilcoxon signed-rank test

The Wilcoxon signed-rank test allowed us to statistically determine whether the mean rank between the data produced with on-the-fly coding and the data produced with post-session coding was null (null hypothesis H0). Since the on-the-fly coding and post-session coding were both used on the same sample we were able to apply the matched Wilcoxon test. When the null hypothesis (H0) was tested, no significant statistical difference was found between the two samples. This meant that the difference between the variables was not null. Variable X was defined as the number of interaction occurrences. This non-parametric test was chosen because the X variable does not require a normal distribution [41] for the paired-sample t-test to be applied. The Wilcoxon test compares the “W_{obs}” observed distribution with the “W_{crit}” theoretical distribution. Based on an error risk of $\alpha < 0.05$, the following results presented in table 7 were obtained.

Condition	W _{obs} (+)	W _{crit} (+)	W _{obs} (-)	W _{crit} (-)	P-values
Stimulo Standard	26	30	10	6	$Z=-1.120, p=0.263$
Stimulo AR	31	30	5	6	$Z=-1.825, p=0.068$
Stimulo SAR	15.5	30	20.5	6	$Z=-0.350, p=0.726$
Artefice Standard	27	24	1	4	$Z=-2.197, p=0.028$
Artefice AR	17	24	11	4	$Z=-0.507, p=0.612$
Artefice SAR	43.5	44	11.5	11	$Z=-1.633, p=0.102$

Table 7: Results of the Wilcoxon rank sum test (validation if $W_{obs} (-) \geq W_{crit} (-)$ or if $W_{obs} (+) \leq W_{crit} (+)$)

Given the results of this statistical test, we were able to conclude that H0 was not rejected (there was no statistical difference between the quantitative data from the on-the-fly coding and the post-session coding): $W_{obs} (-) \geq W_{crit} (-)$ or when $W_{obs} (+) \leq W_{crit} (+)$ and $p > 0.05$.

However, two conditions out of the six listed in the table above failed to validate H0. With respect to Stimulo AR, $W_{obs} (-) \approx W_{crit} (-)$ and $W_{obs} (+) \approx W_{crit} (+)$ and $p = 0.068$. These results clearly did not validate H0. Regarding Artefice Standard, $W_{obs} (-) \not\geq W_{crit} (-)$ or if $W_{obs} (+) \not\leq W_{crit} (+)$ and $p = 0.028$,

hence invalidating H0. It should be noted that the Artefice standard session was very long (1:30) and occurred at the end of the day. The weak results would therefore seem to be due to the coders' waning attention suggesting that the on-the-fly coding results were not fully reliable for this session. This is certainly a factor to be considered in the future.

While the on-the-fly coding method still needs improvement, the statistical test does lend weight to the argument that the quantitative data results are statistically similar to those obtained from standard post-session coding. At this stage, and in answer to our first research question, it can be said that **the on-the-fly coding method does indeed generate data sets comparable to those obtained with a post-session coding method.**

Discussion

In the previous section we presented the internal and external validation of our on-the-fly method, comparing the proportion of end-users' and designers' interactions according to the type of artefact used. The comparison between post-session and on-the-fly coding showed no significant differences in the results for four out of six of the conditions. In this section, we shall discuss the limitations and potential improvements of the tool. We shall also present some results from the analysis of the co-design sessions. These illustrate how original representations can be produced rapidly using the interface. Our aim is to show how the real-time coding method and the *Observer* tool can provide original, fast and relevant data representations allowing primary analysis, as well as facilitate the selection of relevant episodes within large data sets. This will address our second research question: **"what kind of results can be extracted and automatically displayed from on-the-fly coding?"**

4.1 Coding time comparison

The first benefit of the on-the-fly coding method is that it allows a large sample of data from numerous situations to be gathered while remaining frugal in terms of manpower and time. The estimated time savings are shown in table 8 below where the time needed to code a 30-minute protocol using both the on-the-fly and post-session methods is detailed. The ratio observed is approximately 1:2 in favour of our method, which is mainly due to the absence of any post-session coding time. This value is quite far removed from that given by [28] who quote a ratio of 1:10, which includes the time necessary for analysis. The training session for on-the-fly coding is a critical step for coders to reach a satisfactory level of performance in the use of the tool. Even though the training is also mandatory for post-session coding, on-the-fly coding has a lower tolerance to hesitations/errors since it is performed in real time.

Table 8: Time required to complete the coding of a single 30-minute session using the on-the-fly and post-session coding approaches.

<i>Post-session coding</i>	<i>Processing time (min.)</i>	<i>On-the-fly coding</i>	<i>Processing time (min.)</i>
Time to install the recording setting	90	Customising the Observer interface(s)	5
Session recording	30	Training and learning of the coding book and the Observer interface	60
Time to pack up the recording setting	30	Session coding	30
Training and learning of the coding book	60	Automatic and manual merging	120
Video processing	60		
Coding	240		
Total	480	Total	215

4.2 Displaying aggregate results

Researchers can rapidly extract numerical results from the data collected on-the-fly and produce data sheets detailing occurrences, proportions and percentages. The results presented in the three figures below display interesting and contrasting representations. For example, we can see that in the standard condition all the possible interaction modalities are used by the participants (e.g. tangible, digital, ephemeral and none) (figure 11), whereas for SAR and AR conditions the principal artefact is the dominant one (mixed for SAR, figure 9 and digital for AR, figure 10). The diagrams also quickly illustrate that end users are less often the initiators of interactions, whatever the modality. This indicates that the sessions probably remain largely under the control of the designers.

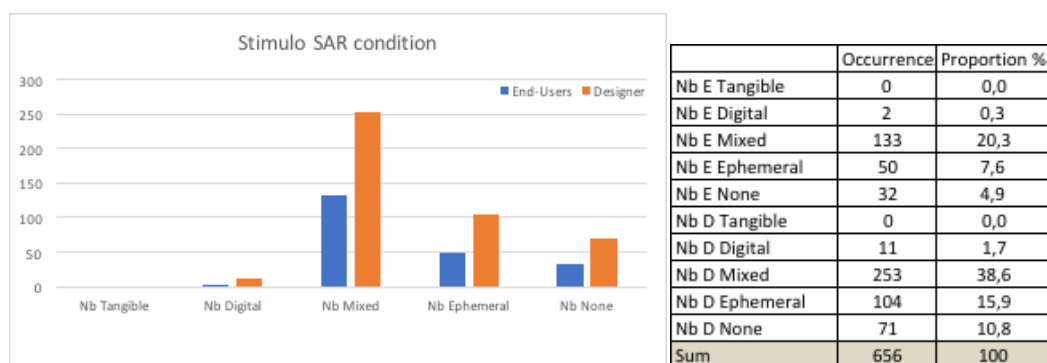


Figure 9: Display of aggregate interactions during the Spatial Augmented Reality co-design session

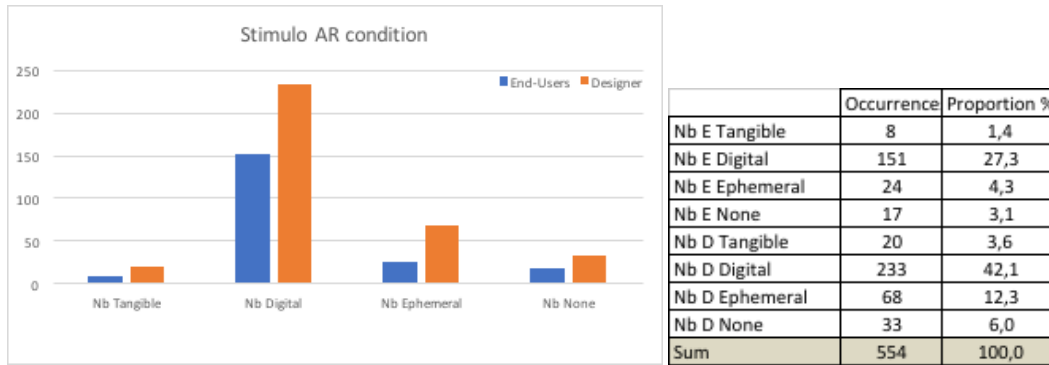


Figure 10: Display of aggregate interactions during the Augmented Reality co-design session

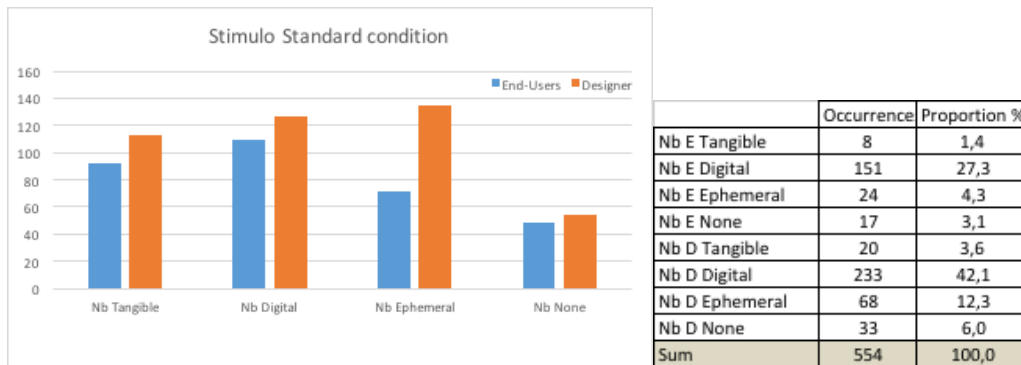


Figure 11: Display of aggregate interactions during the Standard co-design session

4.3 Timelines and interaction densities

A more original set of results can be displayed taking advantage of the temporal dimension of the data set once the data has been merged. Several representations of the results can be provided to researchers and tailored to suit their needs. We present here the most generic temporal representation that can be displayed instantly after the end of each session's merging.

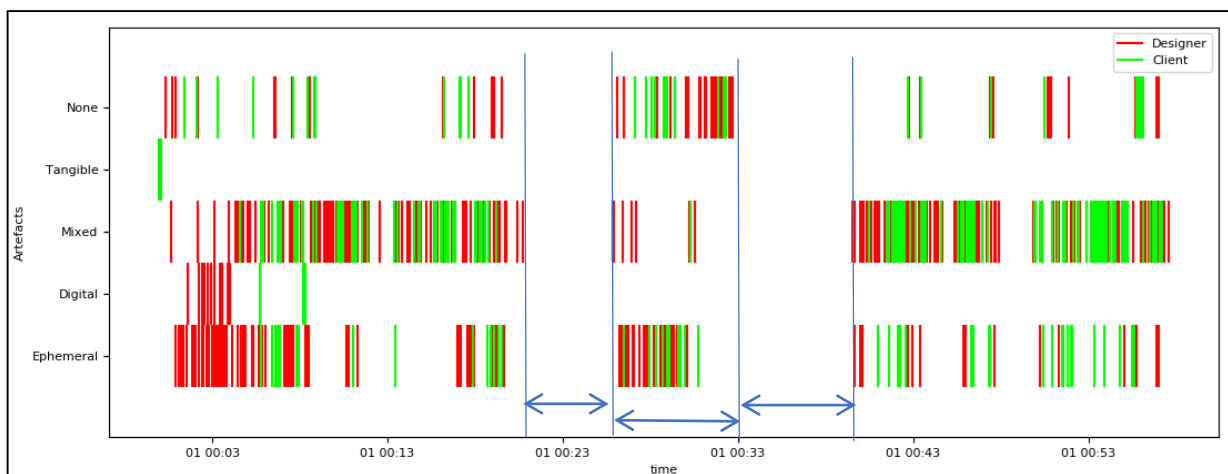


Figure 12: Timeline of the interaction events, excerpt from the Stimulo SAR session



Figure 13: Zooming into the data with the *Observer* pyplot interface (25th to 33rd minute)

Figure 12 shows the timeline of the SAR co-design session outlining the overall profile of the session. In addition to the statistical results discussed previously, the introduction of the temporal dimension increases the session's rapid analysis range. The types of interaction occurring during the session can be identified with a colour code specifying the type of actor (red for designers and green for end users). Such plots are rendered thanks to a module fully integrated into the *Observer* tool and based on a pyplot library (Figure 13). The diagram is displayed inside a box which provides basic functions for its exploration: zoom, pan, etc. It fulfils multiple objectives. The first benefit of this tool is that it rapidly segregates phases of particular importance from those that are not relevant for further analysis. For instance, we can clearly identify two episodes where no artefact-centric interactions happen (between the 21st and 25th minute and between the 33rd and 39th minute, figure 12). Basically, these empty spaces correspond to moments when participants communicated without speaking (which is rare), or worked on their own or experienced technical problems (which was the case here). Moments of intense activity are also easily identified. The researchers can thus focus their attention on these episodes and select them for further processing: video processing, transcription (and translation if necessary). Such interaction profiles also enable the researchers to classify sessions according to their potential relevance. This is especially useful if they wish to study a certain type of artefact-centric interaction for example.

The diagram produced also helps the researchers to visually identify patterns. Specific sequences of interactions are often searched for in design activity analysis with the aim of understanding whether any are recurrent or whether specific session phases rely on specific patterns of interactions. For instance, a researcher might wish to check whether the end users react to a designer's interaction with the same kind of artefact. Session sequences can also be identified. For example, the predominance

of the designer's interactions at the beginning of the session (figure 12) corresponds to the initiation phase where the problem is presented.

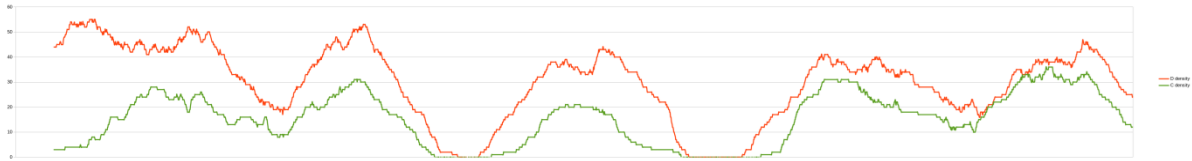


Figure 14: Interaction density over time (i=120 seconds)

In the diagram showing interaction density over time (Figure 14), the curves are computed by adding each interaction performed by the actor concerned in a predefined interval of time l (density of interactions, $d(t) = \sum_{t-i}^{t+i} interaction(t)$). Hence, each point on the diagram at time t represents the number of events counted in the interval $(t-i; t+i)$. This kind of diagram offers another way to rapidly identify moments of intense interaction inside the timeline. The superimposition of the curves makes it possible to analyse interaction dynamics. In this example, it can clearly be seen that interaction density is often much greater for the designers (red curve) than for the end users. This can initially be interpreted as the designers taking the lead in the session. However, all the curves follow a similar pattern, suggesting that the interactions are quite balanced between the parties and that the end users react to the designers' interactions, albeit with a slight delay.

While the value obtained for figure 15 is $i=40$ seconds, the value for figure 14 is $i=120$ seconds. It can thus be seen that higher values of i produce smoother curves. These are preferable for a first overview of the session but may hide a number of odd phenomena occurring in a very short period of time. For instance, figure 14 would seem to indicate that the designers' interaction intensity is always greater than that of the end users. However, a lower i value provides a more detailed view of the situation. For example, specific phenomena come into view such as a change to the interaction intensity hierarchy. Figures 15a and 15b demonstrate a reversal in general behaviour: here the end users interact more than the designers (the green curve oversteps the red one). These moments where the end users are more active than the designers are of particular interest to the researchers. Further analysis should help to understand what causes this shift. In the overall context of co-creative sessions, specific moments of high end-user engagement must be scrutinised. Finally, another important visual clue is the shape of the curves and how they evolve over the course of the session. Figure 15b shows intense activity but, unlike figure 15a, the mean interaction value remains high, suggesting that the interactions are more regularly spread over time.

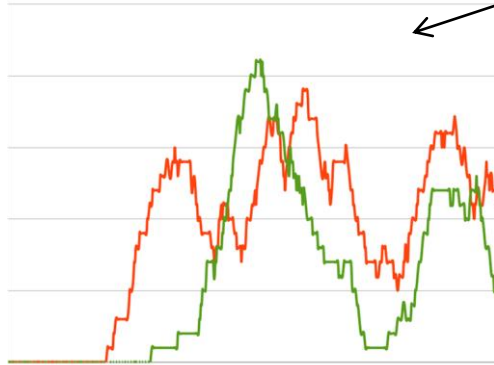
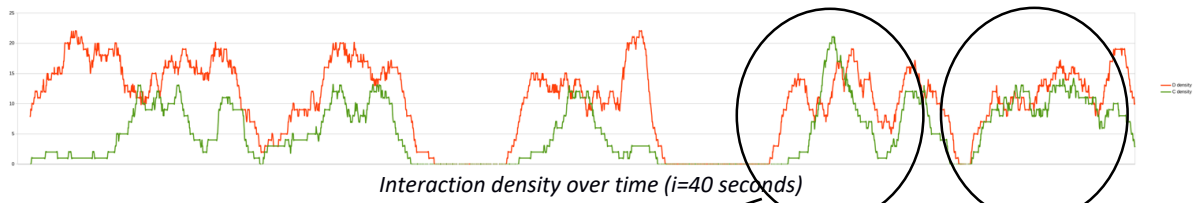


Figure 15a: End users interact more than designers

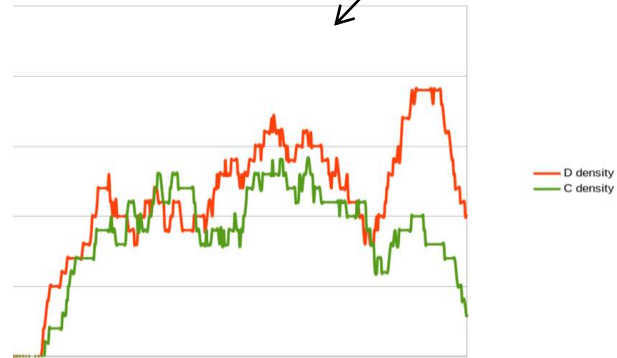


Figure 15b: Zone with sustained intense interactions.

Of course, such diagrams can be produced for any type of event and the curves reflect a specific focus on the types of artefact. Numerous combinations of factors can be used allowing researchers to define the best representation of the phenomenon they are interested in.

The evidence provided in the above sections offers an affirmative answer to our second research question. The *Observer* tool and method allow researchers to automatically display results for interaction distribution and density and to zoom in on different levels of granularity. They highlight interesting information about the session profiles and allow a selective analysis of the more relevant episodes in each session.

4.4 Potential expansion of the method

We have shown in this paper that the on-the-fly method is a fast and reliable technique for coding and processing gesture-based interactions in co-design sessions. Furthermore, we have seen how the analysis of on-the-fly scripts can help to identify relevant design episodes before applying a more in-depth analysis framework.

The observed session configuration has a strong influence on the applicability of the method. High simultaneous interaction rates (usually correlated with the number of participants) and long session durations may flaw the coding. The effects of noise, light, sitting conditions, working hours, etc., are likely to increase the coder's cognitive load and enhance fatigue. One possibility for the future would be to explore the influence of session structure on the coding.

Ergonomic studies could also enhance coder performance by improving the *Observer* tool interface. Several parameters such as the size, shape and location of the buttons might improve coding accuracy and limit the coder's fatigue. An ultimate step could involve developing a specific apparatus for on-the-fly coding such as a handle with physical buttons allowing coders to code without looking at the coding device and thereby reducing cognitive load. In case of complex, intensive or long sessions involving many stakeholders, it might be a good idea to have more than two coders. If the session is long, the coders could work in sequences and if it is dense they could work in parallel. Two coders could also perform simultaneous real-time double-coding of the whole session to improve coding robustness. Testing different coding configurations would provide insights into the optimal number of coders together with the best time and space distribution.

With respect to interactions, studying the duration of their occurrence may also prove to be insightful. This information can already be derived from the actual coding scheme if we assume that there is no interruption between two interactions, i.e. that they follow on from each other. If it is not possible to make this assumption, the end of the interaction could be coded as well as the beginning.

5 Conclusion

This paper presented a method able to provide quick, quantitative data about gestural artefact-centric interactions in co-design sessions. The on-the-fly method put forward was supported by a software tool. The reliability of the method was validated using conventional statistical tests. The data obtained using the on-the-fly and post-session coding methods were compared. The results show that the on-the-fly method is accurate insofar as the quantitative data obtained were close enough to those obtained with a post-session coding method. Statistical tests showed good coherence between the two samples. The results obtained through this method offer valuable information for identifying interesting moments within a session, as highlighted in the discussion session. The method provides interesting and direct access to interaction density, which is an important indicator for identifying potentially interesting episodes. These episodes can then be analysed in more depth using a conventional design protocol analysis. Compared with traditional methods, this on-the-fly approach may well serve as a more efficient and frugal design observation practice requiring significantly less time to complete whilst maintaining a high level of robustness.

Acknowledgment

The work reported in this paper was completed as part of the SPARK project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement

No. 688417. This paper only reflects the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

References

1. Caridà, A., Melia, M., Colurcio, M.: Business model design and value co-creation: Looking for a new pattern. In: Russo-Spena T., Mele C., Nuutinen M. (eds) *Innovating in Practice*. Springer, Cham. (2017)
2. Ardito, C., Buono, P., Costabile, F., Lanzilotti, R., Piccinno, A.: End users as co-designers of their own tools and products. *Journal of Visual Languages and Computing*, Volume 23 Issue 2, pp. 78-90 (2012)
3. Steiner, I. D.: *Group Process and Productivity*. New York, NY: Academic Press (1972) Cited by Kvan, T.: Collaborative design: what is it? *Automation in Construction*, vol. 9, pp. 409-415 (2000)
4. Verlinden, J., Horváth, I. and Nam, T.-J.: Recording augmented reality experiences to capture design reviews, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 3 No. 3, pp. 189–200 (2009)
5. Maurya, S., Arai, K., Moriya, K., Arrighi, P.A. and Mougnot, C.: A mixed reality tool for end-users participation in early creative design tasks. *International Journal on Interactive Design and Manufacturing*, Springer Paris, Vol. 13 No. 1, pp. 163–182 (2019)
6. Schifferstein, H. N. J., & Desmet, P. M. A.: Tools facilitating multi-sensory product design. *Design Journal*, 11(2), 137–158 (2008)
7. Raskar, R., Welch, G., Fuchs, H.: Spatially augmented reality. In: *Proceeding of the First IEEE Workshop on Augmented Reality*; 1998 November 1; San Francisco, CA; pp.63-72 (1998)
8. Park, M.K., Lim, K.J., Seo, M.K., Jung, S.J., Lee, K.H.: Spatial augmented reality for product appearance design evaluation. *Journal of Computational Design and Engineering*. Volume 2, Issue 1, January 2015, pp. 38-46 (2015)
9. Hannah, R., Joshi, S., Summers, J.D.: A user study of interpretability of engineering design representations. *J. Eng. Des.* 23, 443–468 (2012)
10. Ericsson, K. A., & Simon, H. A.: *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press (1993)
11. Cross, N., Christiaans, H., Dorst, K. (eds.): *Analysing Design Activity*. John Wiley & Sons, Chichester, UK (1996)
12. Mondada, L.: Video Recording as the Reflexive Preservation and Configuration of Phenomenal Features for Analysis. In: Knoblauch, H., Raab, J., Soeffner, H.-G., Schnettler, B. (eds.). *Video Analysis*. Bern: Lang (2006)
13. Eastman, C.M.: Cognitive processes and ill-defined problems: a case study from design. In *Proceedings of the First Joint International Conference on Artificial Intelligence*. Bedford, MA: MITRE (1969)

14. Visser, W.: Use of episodic knowledge and information in design problem solving. *Design Studies*, Elsevier, 1995, Analysing Design Activity, 16 (2), pp.171-187 (1995)
15. Lloyd, P.: Can concurrent verbalization reveal design cognition ? *Design Studies* 16, pp. 237-259 (1995)
16. Jaspers, M.W.M., Steen, T., Van den Bos, C., Geenen, M.: The think aloud method: a guide to user interface design. *International Journal of Medical Informatics*. Volume 73, Issues 11–12, November 2004, pp. 781-795(2004)
17. Hay, L., Duffy, A. H. B., McTeague C., Pidgeon L. M., Vuletic T., Grealy, M.: A Systematic Review of Protocol Studies on Conceptual Design Cognition: Design as Search and Exploration. *Design Science* 3 (2017)
18. Gero, J.: Design Prototypes: A Knowledge Representation Schema for Design, *Artificial Intelligence Magazine*, Volume 11 Number 4 (1990)
19. Purcell, A. T., Gero, J. S., Edwards, H., Matka, E.: Design fixation and intelligent design assistants, in J. S. Gero and F. Sudweeks (eds), *Artificial Intelligence in Design '94*, Kluwer, Dordrecht, pp. 483-496 (1994)
20. Amhed, S. & Hansen, C.T.: A decision making model for engineering designers, sahin T.M.M. ed, Computer based design, EDC2002, Cambridge, pp. 217-227 (2002)
21. Girod, M., Elliot, A.C., Wright, I.C., Burns, N.D.: Decision making in conceptual engineering design: an empirical investigation, Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture*, 217 (9), pp. 1215-1228 (2003)
22. Tang, J.C. & Leifer, L.J.: A Framework for Understanding the Workspace Activity of Design Teams, In Proceedings of ACM conference on Computer-supported cooperative work: ACM (1988)
23. Radcliffe, D.F. & Lee, T.Y.: Design methods used by undergraduate engineering students, *Design Studies*, vol. 10, no. 4, pp. 199-207(1989)
24. Olson, G.M. & Olson, J.S.: Small group design meetings: an analysis of collaboration, *Human-Computer Interaction*, Vol 7 pp. 347-374 (1992)
25. Cross, N. & Cross, A.C.: Observations of teamwork and social processes in design, *Design Studies* 16, pp. 145-170 (1995)
26. Wiltschnig, S., Christensen, B.T., Ball, L.J.: Collaborative problem-solution co-evolution in creative design. *Design Studies*, 34 (2013), pp 515-545 (2013)
27. Waldron, M.B. & Waldron, K.J.: Methods of studying mechanical design. In Waldron, M.B. and Waldron., K.J. (Eds.), *Mechanical Design: Theory and Methodology*. New York: Springer-Verlag, pp. 21-34 (1996)
28. Jiang, H. & Yen, C.-C.: Protocol analysis in design research: a review. In *Rigor and Relevance in Design: IASDR 2009*, Seoul, Korea October 18_22, Seoul, pp. 147-156. International Association of Societies of Design Research

29. Perry, M.: Coordinating joint design work: the role of communication and artefacts. *Design Studies*, 19(3), 273–288 (1998)
30. Visser, W.: Function and Form of Gestures in a Collaborative Design Meeting. In S. Kopp & I. Wachsmuth (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction*, Vol. 5934, pp. 61–72, Springer Berlin Heidelberg (2010)
31. Härkki, T., Seitamaa-Hakkarainen, P. & Hakkarainen, K: Hands on design: comparing the use of sketching and gesturing in collaborative designing, *J. Design Research*, Vol. 16, N°. 1, pp. 24-46 (2018)
32. Eris, O., Martelaro, N., Badke-Schaube, P: A comparative analysis of multimodal communication during design sketching in co-located and distributed environments, *Design Studies*, vol.35, pp. 559-592 (2014)
33. McNeill, D.: *Hand and Mind: What gestures reveal about thought*. Chicago: University of Chicago Press (1992)
34. Kang, S., Tversky., B., Black, J., B. (2015). Coordinating gesture, word, and diagram: Explanations for Experts and Novices. *Spatial Cognition & Computation*, 15:1-26.
35. Jeantet, A.: Les objets intermédiaires dans les processus de conception des produits, *Sociologie du travail*, (3), pp. 291-316 (1998)
36. Vinck, D.: Taking intermediary objects and equipping work into account in the study of engineering practices. *Engineering studies*, vol 3, issue 1, pp. 25-44 (2011)
37. Boujut, J.F. & Blanco, E.: Intermediary Objects as a means to foster co-operation in Engineering Design, *Computer Supported Cooperative Work*, vol 12, pp. 205–219 (2003)
38. Becattini, N., Masclat, C., Ben-Guefrache, F., Prudhomme, G., Cascini, G., Dekoninck, E.: Characterisation of a co-creative design session through the analysis of multi-modal interactions. *Proceeding of the International Conference on Engineering Design ICED2017*, pp. 479-488 (2017)
39. Ahmed, S., Wallace, K. M., & Blessing, L. T.: Understanding the differences between how novice and experienced designers approach design tasks. *Research in engineering design*, 14(1), 1-11 (2003)
40. Dorst, K.: Analysing design activity: new directions in protocol analysis, Editorial of special issue of *Design studies*, vol 16, 2, p 139-142 (1995)
41. Shapiro, S.S. & Wilk, M.B.: An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52, 591-611 (1965)