D5.1
VALIDATION AT END USERS' PREMISES

Approval Status

	NAME AND SURNAME	ROLE IN THE PROJECT	PARTNER
AUTHOR(S)	Giuseppe Bellucci	WP leader	ARTEFICE
	Niccolò Becattini	Assistant PC	POLIMI
	Gaetano Cascini	Project Coordinator	POLIMI
	Jamie O'Hare	Deputy PC	Univ. Bath
	Xavi Majoral		STIMULO
	Jean-François Boujut		GINP
	Fatma Ben Guefrache		GINP
REVIEWED BY	Jamie O'Hare	Deputy PC	Univ. Bath
	Niccolò Becattini	Assistant PC	POLIMI
	Gaetano Cascini	Project Coordinator	POLIMI
APPROVED BY	Gaetano Cascini	Project Coordinator	POLIMI

History of Changes

VERSION	DATE	DESCRIPTION OF CHANGES
01	06/06/2018	Structure of the deliverable
02	20/06/2018	Early contributions from partners
03	29/06/2018	Consolidation of the deliverable structure and contents
04	05/07/2018	Results about Artefice Sessions
05	18/07/2018	Results about Stimulo Sessions
06	23/07/2018	First draft of the deliverable (discussion of results and conclusions missing)
07	24/07/2018	Discussion of the results (Section 4.6) v1
08	25/07/2018	Conclusions (Sect 5)
09	26/07/2018	Executive Summary added
10	26/07/2018	Complete draft for PC review
11	30/97/2018	Final revision and copy-editing

Document Details

DISSEMINATION LEVEL	PUBLIC
DUE DATE	30.06.2018 (postponed to 31.07.2018 – approved by the PO)
ISSUE DATE	30.07.2018
CONTRACT NUMBER	H2020-ICT/2015-688417
ELECTRONIC FILE LOCATION	
FILE NAME	D5.1_WP5_Validation at end-users premises

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

This document reports on the activities carried out within the Task 5.1 of the SPARK project, specifically, the testing of the SPARK Spatial Augmented Reality platform at the premises of the consortium's end users. The aim of this activity was as to validate the usability, the effectiveness and the usefulness of the SPARK platform in real operational environments. This activity was completed using the second release of the SPARK platform; the third and final release is expected by the end of M31. The activities of T5.1 were also designed to provide useful evidence to support the fulfilment of Objective #3 of the project proposal:

Study and analyse how and to what extent the SAR technology can stimulate and enhance design creativity through a comparison against pre-defined metrics in real operational design environments.

Moreover, as Objective #4 of the project proposal aims to "*Demonstrate the effectiveness of the SPARK platform in wider real contexts and showcases*", it is worth recalling that the condition of carrying out experiments in real operational environment is also relevant to the initial demonstration of the platform in real contexts.

The document presents the methodology that was applied to organize and run the experiments of T5.1, including the data recording activities, the post-processing, and the analyses required to validate the SPARK platform in real operational environments.
The 'real operational environment' used in these experiments where the premises of Artefice and Stimulo.
The key features of the methodology included:

- The organization of SPARK rooms at end users' premises (arrangement and type of equipment, accessibility to the prototype…);
- The 5 case studies (including the profile of co-designers participating in co-creative design sessions) considered for the experimental activities:
    - 3 cases at Artefice (Design of packaging for three different organic food products);
    - 2 at Stimulo (Product interface design of a medical and an electronic device).
- The metrics used to characterize the co-creative design sessions in terms of:
    - Session performance metrics – which assess the creative outcomes of the sessions;
    - Design process-related metrics – which focus on the gestural interaction of co-designers and on the analysis of the SAR log files (which proved to provide consistent results compared to standard analysis of spoken interactions to unveil design moves);
    - Design process efficiency metrics – which assessed the impact of using the SPARK platform on the overall design, effort and duration of the project.

The results of the experimental activities have been also backed and enriched through interviews with co-designers participating as subjects of the observations during the experiments, so as to gather a more qualitative, but complementary, feedback on the aspects that are difficult to capture through the quantitative metrics adopted within the project. In particular, the interviews focused on the preparation activities with the SPARK platform, before the co-creative design session starts.
The analysis of the experimental data enabled drawing some partial conclusions on the effectiveness of the SPARK platform (and on SAR for design) and the intrinsic efficiency it brings to some of the relevant activities occurring during a design project.
The main findings to emerge from the experiments of Task 5.1 were as follows:

- SAR technology is beneficial in stimulating and enhancing design creativity, as the test sessions completed in T5.1 have shown a number of significant performance improvements in comparison to the results of WP4.
- The use of SAR technology has been shown to provide significant improvements in design process efficiency, according to the first longitudinal case explored in T5.1 (with reference to the data on HR and economic resources consumptions the end users provided about previous non-SPARK-supported projects).
- The SAR appears to enhance the communication between designers and clients/end-users: most of the interactions between co-designers happen through the mixed prototypes (instead of purely digital or ephemeral interactions with artefacts/prototypes).

The experimental activity of T5.1 also highlighted that there are some limitations that should be taken into account for the future exploitation of the project, as these elements might potentially hinder its proficient adoption by stakeholders.

- The session preparation activities (e.g. the preparation UV map and of the assets) should be made simpler or training services should be developed to support the business exploitation;
- Small prototypes might be not properly rendered due to the limitations of the projection technology (e.g. pixel size) and in such case, the SPARK platform should also allow the use of a simple AR scenario (e.g. with a tablet or other devices for visualization).

The SPARK user interface, moreover, has improved in terms of usability, but future developments of the platform, also beyond the timeframe of the project, should focus on the following items:

- The interaction (e.g. file exchange) with software commonly used to develop assets for packaging design (e.g. Adobe Illustrator, Photoshop etc.);
- More precise management of assets when placing them on the prototype (i.e. "stick to grid" option);
- The ability to fine-tune key parameters – such as rotation and scale of assets.

Section 4 of this document presents the further details about these insights and recommendations.

2. INTRODUCTION

This deliverable reports on the results of validation tests at the end users' premises, which are intended to verify the impact of the SPARK Spatial Augmented Reality (SAR) technology on the development of real projects. Topics of design sessions have been selected considering Artefice and Stimulo ongoing projects and were focused on FMCG packaging (Artefice) and digital interaction design (Stimulo). Due to the nature of its business and clients, AMS projects have been considered more suitable for activities related to Task 5.3.

Tests were carried out at end-users' premises under the supervision of Polimi, for both data recording and setting up the SPARK platform (release two) equipment. The SPARK platform has been used to support the evaluation of the creative proposals and the ideation of new possible outcomes – based on the feedback and discussion with the clients during the session.

The results and observations from these activities have been processed with the metrics detailed in T4.2 and compared against the common practice of designers.

The final outcomes and insights contribute to the definition and improvement of the SPARK platform, providing feedback for the development of the platform before the third release (T3.2).

Within this deliverable, section 3 provides the overview of the experimental protocol and the methodology of experimental activities. This includes the details related to real projects observation, a

description of the selected case studies and participants, and a description of the metrics adopted for the analysis and evaluation.
Section 4 focuses on the description of relevant highlights and notes emerged during the sessions, it shows the results of the different analysis made by the academic partners, and it presents an overall discussion about them. Finally, section 5 presents the conclusions of this activity.

3. PROTOCOL AND METHODOLOGY FOR THE EXPERIMENTAL ACTIVITIES

The experimental activities of Task 5.1 focus on the validation of the SPARK platform in real operational environments and therefore they do not include a comparison against different operational conditions, as this was completed within WP4. Nevertheless, the experimental activities of Task 5.1 benefit from the findings and insights from those earlier experiments, particularly the new functionalities that have been integrated into release 2 of the SPARK platform. Furthermore, the type of analysis and the related experimental setting of WP4 showed to be effective in providing interesting data and building meaningful conclusions about the effectiveness of the SPARK platform. The current experiments, therefore, make use of a similar approach in terms of the arrangement of the equipment and method of the analysis.

Among the useful previous findings, the consortium also considered the limiting factors for an effective co-creation during the design session, as emerged during WP4 experiments, such as the placement of co-designers with reference to the position of the mixed prototype and the size of the surface of interaction with the SPARK GUI (improvements presented in section 3.1). Data from the experimental setting got processed consistently with the metrics presented in Deliverable 4.2 (details in section 3.3).

Moreover, it is worth recalling that the longitudinal study initially planned for WP4, which involved the collection of existing data concerning the use of resources in previous end users' projects, for comparison with a case study completed using the SPARK platform. The consortium decided the comparison should be made against experimental data from a real operational environment scenario, to keep consistency. Therefore, this comparison is presented in this document, as the experimental data collected in T5.1 represents an ideal candidate for this purpose.

The case studies used by industrial clients of the consortium end-users are described in section 3.2.

3.1 Organization of the observations

The real operational conditions for the execution of experiments were set-up at the premises of Artefice and Stimulo have been created so that the SPARK platform can work with complete functionalities, in a co-creation scenario that aims at facilitating the interaction, both with the prototype and the media holding the SPARK platform GUI.

Differently from the setting at Polimi and more similarly to GINP's, the prototype SPARK rooms at Artefice and Stimulo have been designed so that a larger number of co-designers have access to the

prototype. Co-designers sit around a round table or by just one side of a squared table, such table in both cases is partially inaccessible as it faces a wall.
Moreover, in these experiments, the interaction media that holds the SPARK GUI is typically larger than a tablet (as used in WP4). This is to allow multiple co-designers to visualize or even interact with the GUI during the session. The multi-touch screen installed at Artefice's premises allowed multiple designers to interact with the GUI, while the mouse on screen scenario used at Stimulo's allowed designers and clients to visualize the GUI.
Figures 01 and 02 respectively show the diagrams of the SPARK rooms at Artefice and Stimulo. Multiple case studies have been tested in each setting (3 at Artefice, 2 at Stimulo) as detailed in the next section.

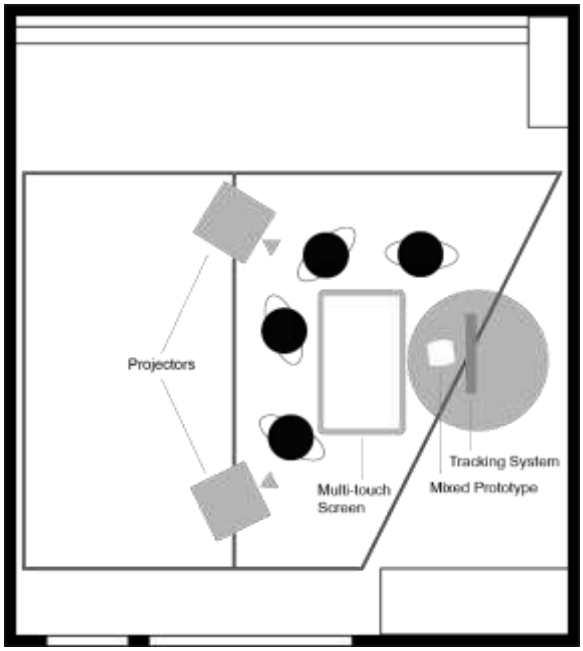


Figure 01: Diagram of the SPARK room at Artefice

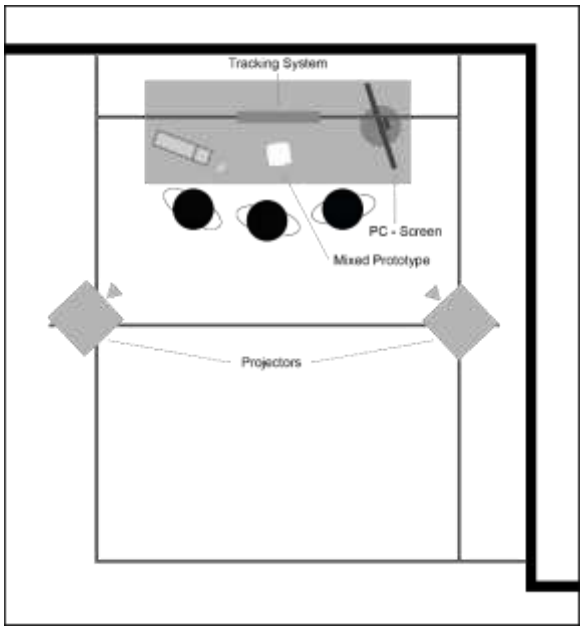


Figure 02: Diagram of the SPARK room at Stimulo

The multiple case studies allow to carry out cross comparisons among sessions of the same kind (packaging design for Artefice, product interface design for Stimulo). Moreover, beyond what emerged in the SPARK project experiments, literature provides no available data about the use of SAR technologies in design. To this purpose, the consortium decided to use WP4 experimental data of the SPARK platform used as a further reference for comparison.

The current setting at both end users' premises allows the external observers to be less intrusive during the session, as they typically stay in the back of the interaction scenario. Moreover, this setting allows the co-designers to focus their attention towards the same direction (or at least to a more limited range). Both these elements concur to arrange the experimental setting as close as possible to a real operational environment (i.e. a scenario that mimics a typical one – not yet supported by the SPARK platform – in terms of accessibility to contents by co-designers and opportunities of design interactions).

Designers completed brief interviews before and after the co-creative design sessions to gather relevant info to:

- Improve the interpretation of design moves during the co-creative process and apply the relevant metrics
- Gather a qualitative feedback about the preparation of the sessions, the reuse of knowledge generated along the sessions as well as changes in the typical workflow of Artefice and Stimulo.

Audio and video data were recorded during the sessions by means of conveniently placed cameras and microphones. Log files were also captured from the system to provide data concerning the usage of the SPARK platform GUI. Further details of the data processing procedures are presented in section 3.3.

3.2 Case studies and participants

This section describes the observation of real-life works, therefore case studies were selected among current projects (at the time of the experiments) from the two end-users companies, according to their calendar and deadlines.
For Artefice Group three design sessions were planned through the month of May 2018. The participating designers were not involved in the previous stages of the project, nor had they had the chance to use the SPARK platform before.
Every session included the following figures:

- Creative Director;
- Designer;
- Account;
- Two Clients, from the Marketing Department.

The three sessions entailed three different projects belonging to the same brand:

a. **Fresh Pizza**: this session happened at the early stage of the Ideas Production phase and in particular the first presentation to the client. The starting point was the brief, from which Artefice Group developed three creative proposals. The session was created in order to share these proposals, having immediate feedbacks and a live active interaction with the client.

The session was structured as follows:

1. Presentation of the three creative proposals to the client, with the collection of feedbacks, impressions and comments.
2. Definition, together with the client, of the best proposal to be developed.
3. Co-creation session: usage and combination of all the graphic elements prepared for the creative proposals, in order to develop potential versions and potential next steps elaboration. The main goal of this session was to shorten the number of potential debrief needed for the project, through the co-creation.

b. Parmigiano Reggiano: the second session also happened at the early stage of the Ideas Production phase and in particular the first presentation to the client. The starting point was the brief, from which Artefice Group developed five creative proposals. The session was created in order to share these proposals, having immediate feedbacks and a live active interaction with the client.

The session was structured as follows:

1. Presentation of the five creative proposals to the client, with the collection of feedbacks, impressions and comments.
2. Definition, together with the client, of the best proposal to further develop.
3. Co-creation session: usage and combination of all the graphic elements prepared for the creative proposals, in order to develop potential versions and potential next steps elaboration.

The main goal of this session was to shorten the number of potential debrief sessions needed for the project, through the co-creation.

c. Mozzarella: the third session still happened during the Ideas Production phase, but this time it entailed the second presentation to the client. It came after the presentation of different proposals, after the initial choices made by the client and the feedback on how to proceed.

The session was structured as follows:

1. Presentation of the debrief, i.e. the re-worked proposal, according to clients feedbacks.
2. Co-creation session: further changes and adjustment were discussed and tested in real-time between the end-users and the client.

Key considerations were:

- The budgeted 'creative' hours;
- The creative brief given by the client;
- The different Stock Keeping Units (SKUs) on which to work.

The Fresh Pizza project was also selected as a 'longitudinal case study' for the application of the design process efficiency metrics. This was considered a perfect benchmark to be followed from start to finish, due to its similarities to the Frozen Pizza project, which was captured as a longitudinal case study as part of T4.2.

This project entailed the creation of a Pack System, with two SKUs and the application of the predefined Brand Visual Identity System. The starting brief of the project was received by the designers on May 10$^{th.}$ and the first presentation to the client happened on May 17th (during the T5.1 session at Artefice premises). Then the project reached, during the first week of June, the last step of the Idea Development Phase.

For Stimulo, two design sessions were planned on July 11th. The participating designers were not involved in the previous stages of the project, nor had they had the chance to use the SPARK platform before. Every session included the following figures:

- Creative Director;
- Designer;
- Two Clients, CEO and an engineer.

The two sessions are based on projects from startup companies: EMZER and NEW BORN SOLUTIONS. Further details of the projects are provided hereafter.

Project: LISN (Line Impedance Stabilization Network)
Client: EMZER
Background: Electrical and electronic equipment need CE certification to be marketed in Europe, which involves compliance with the rules described in the Electromagnetic Compatibility Directive 2014/30/EU. The equipment proposed by EMZER Technological solutions, SL, will assess the compliance with the regulations and will help to eliminate the interferences generated by the equipment under test.

During the first project stage, Stimulo did an exploration of design proposals that might fulfil the project requirements, which included requirements concerning:

- Ergonomics;
- Usability;
- Aesthetics and form;
- Architecture of assembly, assembly and sizing.

From the 2 initial design proposals, Stimulo developed one proposal. During the experimental session, Stimulo needed to validate with the client the best configuration of assets in order to improve usability.

From Stimulo's point of view, it was important to hold this session because the client has a very good technical background but minimal experience in the design of the User Experience in their devices. So, with one design direction selected, the focus of the session was to refine the design for the front and back cover, considering both technical and aesthetic requirements.

The session was structured as follows:

1. Presentation of the final prototype for the main body.
2. Co-creation session: usage and combination of all the graphic elements prepared for the creative proposals, in order to develop potential versions and potential next steps in the elaboration of the product. The main goal of this session was to shorten the number of iterations needed for the project through co-design.

Project: NEOSONICS
Client: NEW BORN SOLUTIONS
The Neosonics is the first non-invasive screening device for infant meningitis. The device employs a sophisticated precision-engineering system, using high-frequency ultrasound to non-invasively count white blood cell in the cerebrospinal fluid below the infant fontanel (the region of the head where the bones are not yet closed).

Stimulo is starting to move from the traditional 'waterfall' design process to a more 'agile' methodology, especially in collaborations with startups. Hence, Stimulo proposed to check and validate the MVP (minimum viable product) using a basic prototype.

The session was structured as follows:

1. Presentation of the assets proposals to the client, with the collection of feedbacks, impressions and comments.
2. Definition, together with the client, of the best proposal to be developed.
3. Co-creation session: usage and combination of all the graphic elements prepared for the creative proposals, in order to develop potential versions and potential next steps elaboration. The main goal of this session was to shorten the number of iterations needed for the project through co-design.

3.3 Evaluation means: performance metrics, usability metrics and follow-up survey

3.3.1 Session preparation feedback

Designers have been asked before the co-creative design sessions (or after when this was their only availability) to provide a qualitative feedback about the use of the SPARK platform. This feedback is gathered by semi-structured interviews, whose topics/questions explored the preparation of the session as well as the changes the use of the SPARK platform introduces in the end users' typical workflow. The topics to be explored have been proposed with questions formulated similarly to the following ones:

- How did you experience the preparation of the SPARK co-creative sessions? Consider the whole activity, from the preparation of assets to the preparation of the sessions on the platform.
- How much did you think this has affected your standard way of doing your job? What would you improve?
- Can you provide an estimation and/or a qualitative description of the benefits of using the platform that you perceived?

The feedback was recorded for further processing (identification of similarities and peculiar differences between scenarios and conditions).

3.3.2 Session performance metrics

Within WP1 a suite of metrics was developed to evaluate the performance of co-creative design sessions. These metrics, and the procedure for their application were subsequently refined and applied within the experimental sessions completed within WP4. The metrics are summarised in Table 1.

Application of the metrics involved a number of data gathering activities. First, a pre-session interview was conducted with the lead designer, just before the start of the session. The designer was asked about their objectives for the session, what ideas had previously been generated in the project, any open tasks from previous sessions (for the Task Progress metric), and how many ideas they would like to end up with by the end of the session (for the Filtering Effectiveness metric).

At the start of the co-creative session itself, all participants were asked to request a screenshot/ picture to be taken of the prototype whenever they felt that they had generated 'a new idea'. To avoid too much disruption to the session, the screenshots/pictures were taken by the designers, either using the

screenshot feature built into the SPARK platform user interface, or by taking a photograph with their phone camera during the 'standard' sessions.

Table 1: The co-creative performance metrics to be applied in WP5.

Metric title	Metric definition
Quantity of ideas	Quantity of ideas generated during the session counted as the number of screenshots taken by participants, verified in the post-session interview.
Variety of ideas	Variety (coverage) - Number of original feature rows that contain a new idea counted on the morphological chart created by a native-speaker observer in the session.
	Variety (new rows) - Number of new feature rows added counted on the morphological chart created by a native-speaker observer in the session.
Quality of ideas	Number of new ideas generated that are taken forward at the end of the session for further development. Determined by participants' consensus in the post-session interview.
Novelty of ideas	Novelty score from 1 to 10 for each of the ideas captured as a screenshot during the session. Determined by participants' consensus in the post-session interview.
Task Progress	Task Progress = 3pts x (Number of high importance tasks resolved or new tasks created) + 2pts x (Number of medium importance tasks resolved or created) + 1pt x (Number of low importance tasks resolved or created). Captured from pre- and post-interview with the session leader.
Filtering Effectiveness	Filtering Effectiveness = Number of ideas rejected ÷ (Number of ideas considered - Desired number of ideas to retain)
Usability	Creativity Support Index survey score

A researcher was sat in the room during each of the sessions to take live notes about the ideas that were being discussed. These notes were captured in the form of a table, based on a 'Morphological Chart', where the rows describe the feature or function that is being discussed, and a description of the potential embodiment options being discussed are captured in the columns. An example of this type of table is provided in Table 2.

Table 2: Example of the 'Morphological Chart'-type table used to for the Variety metric assessment

	Option 1	Option 2	Option 3	Option 4
Tomato graphic position	On right hand side	On left hand side		
Text: Polpa di pomodoro	In Italian	In English		
Italian map graphic	Small	Large		
Text: Picked and preserved in Italy	Picked and preserved in Italy	Picked and preserved within 10 hours		
Text: 100% Italian tomatoes	100% Italian tomatoes	On-plant ripened tomatoes	100% made in Italy	Grown and made in Italy
Text position: 100% Italian tomatoes	Around circle	In red quadrant		
Number of quadrants with text	4	3		

After the session, a joint interview with both the designers involved in the session was completed. In this session, the designers were presented with the screenshots/pictures of the ideas that had been captured during the session. They were asked to confirm that all the ideas had been captured and that none of the ideas was a duplicate or captured by mistake (for the Quantity metric). The designers were then asked to rate each of the ideas in terms of their novelty on a scale from one (low novelty) to 10 (high novelty) (for the Novelty metric) and then asked to decide if each idea would be taken forward in the project for further development (for the Quality and Filtering Effectiveness metrics).
Next, they were then presented with the Morphological Chart that had been captured by the researcher and asked to confirm the accuracy of the chart. They were also asked to identify which of the rows, if any, described new features/functions of the product that had not previously been considered within the project (for the Variety metric). Finally, the list of open tasks from the pre-session interview was revisited to check which tasks had been completed and what new tasks, if any, had been generated during the session.

The usability assessment was completed through the application of the 'Creativity Support Index' (CSI) survey (Cherry & Latulipe, 2014). The CSI tool has been developed based on the NASA TLX survey method, but with a greater emphasis on evaluating creativity support tools (Cherry & Latulipe, 2014). The CSI survey consists of two parts. In the first part, the user rates their level of agreement with 12 statements that cover six aspects of tool performance: collaboration, enjoyment, exploration, expressiveness, immersion, and achieving results that are worth the effort they put into using the tool. In the second part of the survey, the user completes a pairwise comparison of the importance of the six aspects of tool performance listed above. The result of the pairwise comparison is used to generate

weighting factors that are applied to the scores from the first part of the survey. Finally, a score is generated ranging from zero to 100, where 100 indicates that the tool used provided excellent creativity support for the task completed. After each session, all designers that had been involved in the session were asked to complete the CSI survey for the session they had just completed.

3.3.3 Artefact based interaction analysis

As we put forward in WP1, the interaction with artefacts that occur in design sessions is a phenomenon that has been studied in the scientific literature through a protocol analysis method based on the analysis of verbalisations and/or the analysis of gestures associated with verbalisations. Considering the interactions based on gestures is an important level of analysis which can help us to better understand what happens during a collaborative session, and more specifically, a collaborative session using the SPARK platform.

This approach allows us to measure the effect of the SAR on the participation and implication of the end-users' performance during a collaborative session. The analysis framework that we recall here is based on the capture of gestures interaction between the participants and the materials used during the session (including physical prototypes, digital, etc.).

Due to a lack of quantitative studies of design gestures, we found that it was necessary to develop a method for design gesture analysis that could provide a high level of repeatability and robustness. We created our own analysis coding scheme that we tested during three design sessions with professional design teams (Becattini et al., 2017). It is based on three elements: the client(s), the designer(s) and the artefact(s) used to support their interactions with two groups of interactions: with artefacts (Tangible, Digital, Mixed and Ephemeral) and without artefact (None).

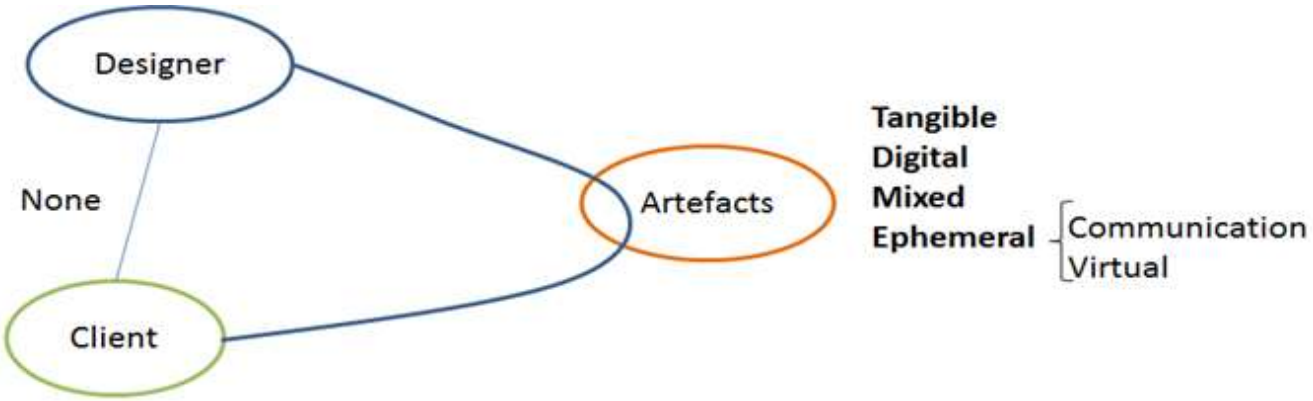


Figure 03. Conceptual framework for the design gesture analysis coding scheme.

Each interaction supported by an artefact was subdivided into different categories: digital, tangible, mixed and ephemeral. In this context, the Digital artefacts category (Figure 3) includes any kind of representation displayed on a screen, laptop, and the Tangible artefact category includes physical mock-ups, printed sheets of the alternatives drawings, sticky-notes that will be posted on the whiteboard, or personal notes. The Mixed artefact category is defined as a physical prototype (physical mock-up with a predefined shape, mostly 3D printed) on which digital elements are projected. They can be pictures, images, text or textures. This is the kind of artefact which are displayed by the SAR (Spatial Augmented Reality) system. 'Ephemeral' artefacts include both 'virtual artefacts' and 'communication gestures'. A virtual artefact is defined as an imaginary object that is depicted or mimicked by a gesture in the air. Communication gestures correspond to gesturing with hands instead of or while speaking. These hand gestures, or 'beat gesturing' (Eris, 2014), complement speech and are not task specific with some symbolic gestures.

In WP5 we needed a scalable research method that could be applied quickly and efficiently to a variety of design sessions. We have therefore designed and validated a method that allows capturing and rapidly delivering data from design sessions showing the relative involvement of designers and clients and the distribution of the type of interactions.
The objective is to reach a quantitative description of gestural, artifact-centric interactions occurrences made by designers and clients during a specific co-creative design session. These interactions are captured only when they occur simultaneously with a spoken utterance - talking is our prompt to record a collaborative activity.

Our method is divided into two steps:
Coding step - the objective of the first coding step is to obtain primary data, doing a coding on the fly during the session. This is done through direct observation with two coders being present during the session.

Analysis step - the second one is to process and analyse these primary data with the objective to get a quantitative description of the interaction occurrences of the sessions.

The coding step has to be made in real time by two coders using a tool we named 'Observer'. One coder has to code who intervenes, more precisely, who is at the origin of the collaborative interaction – a Client or Designer? The second coder has to code the types of artefact that are used to support the interaction between the clients and designers. Interfaces proposed to coders are described figure 04a and 04b. A coding book (including rules for coding actors and artefacts) has been developed, which can be used as a support for training novice coders or as a reminder before each new session for more expert coders.

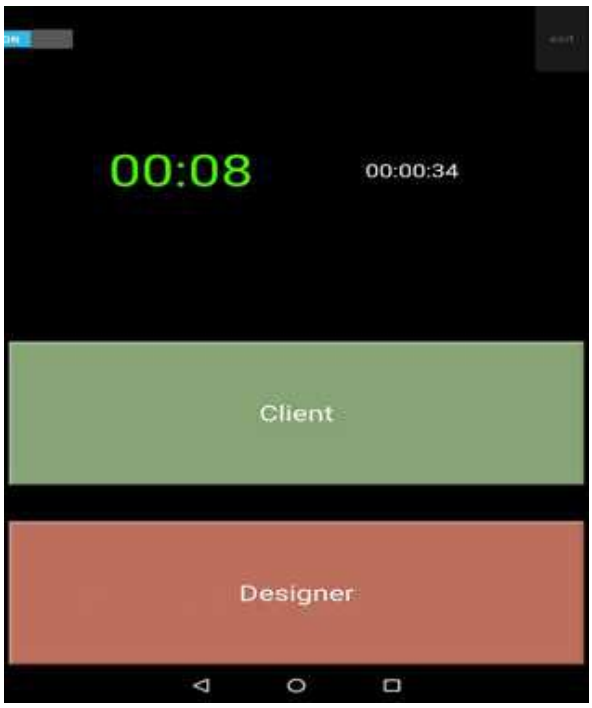


Figure 4a. Interface for coding actors during interactions

Figure 4b. Interface for coding the artefact supporting interactions

3.3.4 Elaboration of the log file of the SAR module

As briefly mentioned in Section 3.1, the verbal interactions taking place during a design session are crucial to gather insights characterizing the ideation dynamics (i.e. the inherent cognition) among co-designers during the design process. Nevertheless, the thorough investigation of these aspects mostly deals with the design features (items and parameters) to be displayed on top of the mixed prototype, being it the primary medium for sharing ideas during the co-design session.

Previous investigations carried out within the consortium for project purposes highlighted that there is a strong positive correlation between what is discussed verbally (according to design protocol analysis) and what is reported as design moves within the log files recorded during the use of ICT alternatives (figure 5, Master thesis of Nicolas Carbone - POLIMI, 2018). That study showed that the log files proficiently captures the main items and parameters characterizing the mixed prototype and their content. Moreover, it has a strong positive correlation with both the main cognitive activities recorded in design protocol analysis (results showed that in all the WP4 co-creative sessions there was a significant match between the items and parameters discussed with design moves of synthesis → solution space and with design moves of analysis → problem space, suggesting an effective, co-evolutionary dynamic as for Dorst and Cross, 2001).

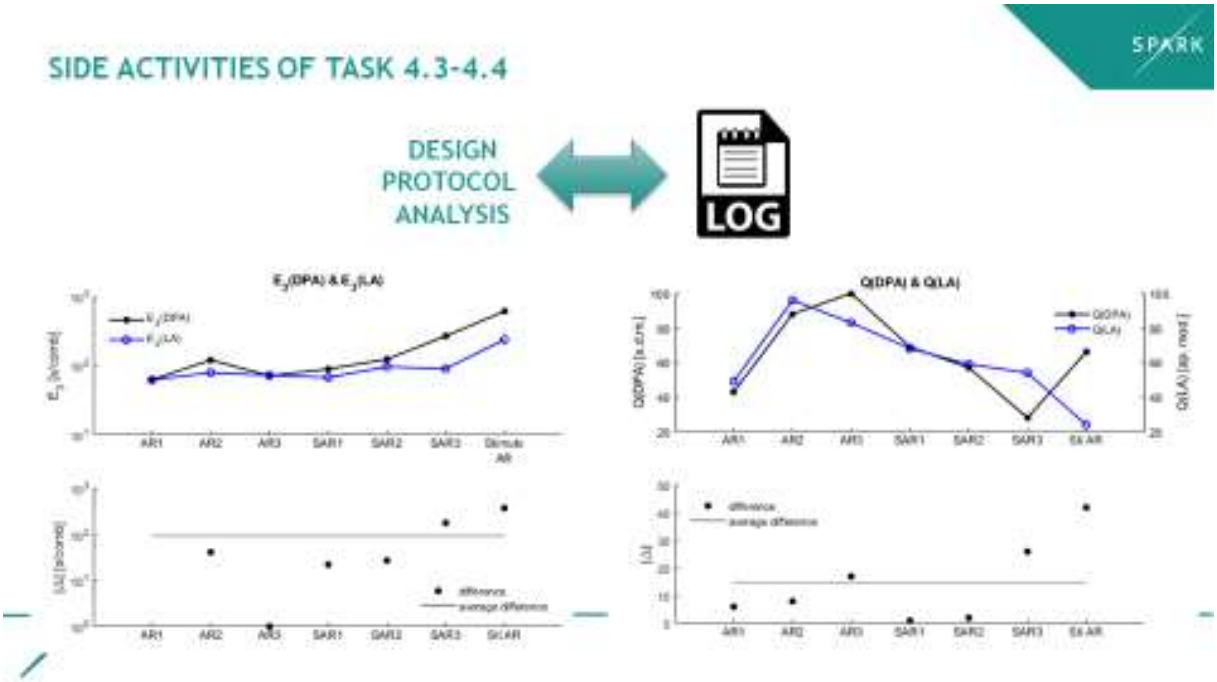


Figure 05 Representative picture showing the comparison between design protocol analysis and log-based protocol analysis for the same design process metrics.

As Task 5.1 does not aim at comparing different technologies (e.g. AR vs SAR), the investigation of the SPARK platform capabilities in these design sessions will focus exclusively on the items and parameters to be displayed and modified onto the mixed prototype. Such framed investigation allows focusing on the interaction co-designers have with the platform, i.e. one of the most critical aspect to improve as emerged after the experiments of WP4 and it is still suitable for the application of relevant metrics to describe co-designers' behaviour in co-creative sessions.

The analysis, based on what was recorded in the log files of the SPARK platform, reused the coding schemes already presented in relevant deliverables from WP1 and WP4.
Consistently with the structure of the log file, which records the elements used for projection and the changes they undergo during the whole co-design session, the coding schemes for items and parameters have been also reduced to a smaller amount of codes. Nevertheless, they proved to be sufficient to fully describe the same dynamics captured in WP4 experiments. Tables 3 and 4 present the coding schemes respectively used for items and parameters.

Table 3: Coding scheme for items, as used for the analysis of log files of co-design sessions

Text	What is expressed by words
Image	A computer-generated picture (potentially vectorial)
Photograph	A photograph of a real object (non-vectorial image)
Icon	A graphic or textual representation related to marks of certification or similar
Background Motif	A texture or a set of elements characterizing the background of the design
System Parts	A reference to part of the entire system
Whole	A reference to the design proposal as a single entity

Table 4: Coding scheme for parameters, as used for the analysis of log files of co-design sessions

Position	The parameter refers to the geographical location of the item on the design proposal or on one of its parts (i.e. [x,y] coordinates)
Orientation	The parameter refers to the degree of rotation of the item (i.e. θ coordinates)
Size	The parameter of the item describes refers to a change in its dimensions (without changing the aspect ratio)
Number	The parameter refers to the quantity of items of the same kind to be added on or removed from the design proposal
Presence	The parameter refers to the introduction or the removal of an item within the design proposal
Colour	The parameter refers to the chromatic characteristics of an item

The analysis of the log files by means of this coding scheme allows the gathering of data about the couples of items and parameters that the co-designers dealt with during the sessions.

The metrics used for the investigation is consistent with those previously used in the SPARK project for verbal interactions, despite the fact that it focuses on the characteristics that can be appreciated independently from the intentions of the designers (as this last facet of the design process has not been coded, as specified above). Data collected in this way can be visually represented as a heat map, in which

the colour of the cells becomes more intense for larger numbers (typically a frequency or number of occurrences).

All the sessions have been measured on a time basis, as the log file records the start and end time of each session. The index Duration measures the session in seconds as the difference end-start.

The Exploration indexes (E1, E2, E3) provided some interesting results in WP4 about the amount of exploration activities (approximate amount of solution variants) the co-designers consider along the sessions. Exploration is computed as 3 slightly different indexes, each of them focusing on different facets of the process and the interaction with the platform.

$$E_1(LA) = \frac{number\ of\ "item-parameter"\ combinations}{maximum\ number\ of\ "item-parameter"\ combinations}$$

E1 provides a measure of "what is used during the session and how", with reference to the whole set of possible item types and the related transformation to those items as allowed by the SPARK platform, therefore it is expressed as a ratio.

$$E_2(LA) = \frac{number\ of\ "item-parameter"\ combinations}{possible\ "item-parameter"\ combinations\ per\ item\ used}$$

E2 provides a measure of "what is used during the session and how" with reference to the whole set of possible item types prepared by the designers for the session together with the related transformation those items can undergo. As for E1, it is expressed as a ratio. E2, therefore, provides an approximate (*) measure of the degree of utilization of the previous work carried out by designers to prepare the SPARK session. (*) As different assets prepared for the sessions get classified within the same type of "item".

$$E_3(LA) = \frac{\Delta time\ block}{number\ of\ "item\text{-}parameter"\ combinations\ per\ Synthesis}$$

E3 is an indirect measure of persistence interacting with items, which was meaningful to compare different testing conditions in WP4(**). It is computed as the ratio between the duration of the session and "what is used during the session and how (explored items-parameters couples).

The Quantity of ideas (Q) is computed as the amount of changes carried out during the session with the assets managed by the SPARK platform. It accounts for everything recorded in the log that reflects an intention of changing the mixed prototype.

$$Q(LA) = number\ of\ applied\ modificaitons$$

To reduce the bias of the variability across the sessions considered, a relative measure of quantity can be computed with reference to the duration of the sessions themselves (Q/duration).

The index V displays the intensity of the variations applied with reference to the overall number of applied modifications (it is expressed as the ratio of the numerator of E1 and Q). A lower ratio corresponds to higher occurrences of applied modifications. While Q increases by one unit every design move recorded in the log file, the numerator increases by one unit only when a new kind of item-parameter couple is explored (e.g. if an image has already been rotated, the rotation of another image does not increase the numerator, but increases the denominator.

$$V(LA) = \frac{number\ of\ "item-parameter"\ combinations}{number\ of\ applied\ modifications}$$

3.3.5 Design process efficiency metrics

As well as assessing the impact of the SPARK platform on the performance of individual co-creative design sessions, one of the ambitions of the project was to evaluate the impact of the technology on the overall efficiency of the design process. The underlying assumption is that the use of the SPARK platform will improve the efficiency and speed of the design process by significantly reducing the number of design iterations and prototypes required to arrive at a suitable final design. To assess the validity of this assumption, a set of design process efficiency metrics were defined in deliverable D4.1 and are summarised in Table 5.

Table 5: Summary of the design process efficiency metrics.

Metric title	Definition
Person-hours spent on the project	All hours spent on the project by design agency (including unbilled hours)
Lead time	Number of days between project start date and end of the Ideas Development phase
Total development cost	Direct costs incurred by design agency (Only up to End of layout - Ignoring post-production costs)
Cost of prototype production	Cost of preparing all design representations used in collaborative sessions or sent to end-user (materials and labour)
Re-work iterations	• Total number of times that a design activity has to be completed again (as requested by management). • Total number of all major design meetings within each design stage • Total number of co-creative design sessions completed within each design stage (and no. of those of which are conducted with the SPARK platform) • Number of versions released to the end-user, feedback received and acted upon

Within T5.1, these metrics were applied to a project that was completed by Artefice for one of their regular customers, who we shall refer to as 'Food Inc.' to maintain the anonymity of the client. The project involved the development of packaging designs for a fresh pizza product with two flavour varieties. Within WP4, benchmark data for the design process efficiency metrics were collected by analysing three projects

completed by Artefice for Food Inc. in which the SPARK platform had not been used. One of these benchmark projects concerned the development of packaging for a frozen pizza project, which was therefore a very relevant benchmark for comparison.
The project was monitored by the Artefice Client Manager for Food Inc., who was able to provide all the relevant data about the project based on the project records and Artefice staff timesheets.

4. RESULT AND DISCUSSION

4.1 Overview of the experimental sessions

Sessions in Artefice, from the end-user point of view, were really successful with a positive impact both on project development and client relation with the agency. Feedback gained from the sessions were really useful to define in a short time, the creative proposal to be finalized for the packaging. Beside this, it is important to underline that after few steps of rework, subsequent to the SPARK sessions, Mozzarella and Parmigiano projects were placed on 'standby' due to internal factors at the client, not related to packing design, but to the products itself. Another important outcome from Artefice sessions relates to some feedback provided by designers involved in these projects. During conversations and preliminary interviews immediately after the sessions, the designers expressed that they were really happy about the impact of the platform in facilitating the exchange of ideas with clients and in the definition of a clear and shared creative proposal at the end of the session.
There were some interesting suggestions about the session preparation process and their indications on some points of that process that could improve the efficiency (considering their specific tasks) and shorten the time effort to dedicate to the preparation. Considering these observations and notes, it was agreed to carry on further tests in the second semester of 2018 (from July to October).

These experimental tests were the first done at Stimulo premises, and for this reason, the comments that emerged during the various stages of the preparation and delivery of the sessions are very interesting.

Hereafter, we report some relevant considerations recorded at Stimulo after this first experience:

- **Pre-calibration**

"It took a long time to set up everything and Stimulo needed further online support from Milano team. This time was considered as a training test for Miquel Pons, chief engineer, that will be in charge of this operation in the future."

- **Calibration**

"During the session, unexpected situations happen and it's hard to modify or adjust the Information System (IS) to fix and carry out the session. This situation may affect the meeting since we couldn't use the tablet and work directly from the IS. That means that some functions were limited, like rotate the assets. We are more worried about this situation than pre-calibration."

- **Assets download**

"Open the session and download the assets was not really a big issue since designers started to work keeping in mind the SPARK platform."

- **Session**

"We ran the session with one single projector (despite having two installed). This is an important limitation since we couldn't "play" with the prototype. To fully explore all the potential of the SPARK platform, we consider it important to have full tracking of the prototype since the session is:

  - *more agile;*
  - *natural;*
  - *engaging.*

During the session, the designer suggested to control/adjust the contrast of the projection".

- **Prototype**

"We really faced a limitation when we are working with a small prototype, especially if we want to interact with clients to improve the User Experience. The small size also affected the assets shapes (from circular to oval)".

Beside the SPARK sessions, Stimulo held two independent sessions with clients:

- EMZER, after the SPARK meeting they had a further discussion regarding manufacturing issues to consider following prototype study.
- NBS, preliminary meeting with a Stimulo chief engineer to have a further discussion regarding mechatronic solutions (a key value point of the project at this stage).

4.2 Feedback gathered through interviews with designers

The interviews with designers carried out after the co-creative design sessions allowed the exploration of several aspects that emerged as potentially critical for the development of the SPARK platform as well as for its exploitation. As mentioned in section 3.3, the interviews aimed at checking the overall experience (strength and weaknesses) of the preparation of a SPARK session, which is a starting point to also explore how the introduction of the SPARK platform affects their typical workflow.

These interviews have been carried out both with Artefice's (4) and Stimulo's (2) designers, all of them involved with different roles in the experimental activities of WP4 and/or WP5. For what concerns Artefice, 2 out of the 4 interviewed designers have been involved in the experiments of both the WPs. The other 2 designers from Artefice got involved in just one of the two WPs (one within WP4, one WP5). For what concerns Stimulo, the interviews were carried out with subjects that were involved just in the experiments of WP5. The following table 6 summarizes their profile and their involvement in the experimental activities.

Table 6: Profile of involved interviewees and their involvement in the experimental activities of the project

Company	Subject (sex)	WPs	Activities carried out within the experimental activities
Artefice	1 (F)	4, 5	• Preparation of assets for co-creative sessions • Preparation of the session with the Information System • Facilitation of the co-creative session as creative director • Interaction with the SPARK GUI (multi-touch screen) • Wrap up of the co-creative sessions and re-elaboration of creative proposals

Artefice	2 (F)	4, 5	• Preparation of assets for co-creative sessions • Preparation of the session with the Information System • Facilitation of the co-creative session as creative director • Interaction with the SPARK GUI (multi-touch screen) • Wrap up of the co-creative sessions and re-elaboration of creative proposals
Artefice	3 (F)	5	• Preparation of assets for co-creative sessions • Preparation of the session with the Information System
Artefice	4 (M)	4	• Preparation of assets for co-creative sessions • Preparation of the session with the Information System • Preparation of the prototype (3D model, UV map…) • Interaction with the SPARK GUI (multi-touch screen)
Stimulo	5 (M)	5	• Preparation of the session with the Information System • Preparation of the prototype (3D model, UV map…) • Interaction with the SPARK GUI (multi-touch screen)
Stimulo	6 (M)	5	• Preparation of assets for co-creative sessions • Facilitation of the co-creative session as creative director

This variety of subjects is considered sufficient to gather a relevant and meaningful viewpoint on the main topics to be investigated with the mentioned questions.

Instead of detailing the answers of each of the above interviewees, the following points summarize the most commonly mentioned opinions among respondents, where relevant peculiar differences characterizing the two design domains (packaging and product) will be highlighted.

The following statement (or similar) have been recorded in almost all the interviews with the 6 subjects.

"*Our workflow is typically different, we need to change our way of working if we have to deal with SPARK*".

Nevertheless, this opinion has similarities and differences between the two domains. For what concerns the packaging design domain, Artefice's designers justified this perceived shift in the workflow from two different perspectives. On the one hand, they consider the SPARK workflow different in essence due to technological reasons. In fact, the contents to be prepared for the projection have to be typically less featured than those they usually prepare for a real-like computer rendering. The designers felt that "fully-featured" contents provide a bad overall visualization on the mixed prototype, which looks more realistic when the asset for the projection has a simpler set of features (e.g. no 3D effects, flat contents…).
On the other hand, the Artefice's designers also recognized that the SPARK workflow makes their way of working also different from the perspective of the whole process of concept development. Typically, one of their designers ideates and creates variants and alternatives of the whole packaging composition for brand identity communication. Later on, these variants are discussed and evaluated with suggestions for improvements in a design review session, which reflect the Artefice's perception and understanding of co-design session, which is different from the expected co-creative design session. Such a difference in meaning is probably also part of the misalignment and the general perception that to prepare a SPARK

session they feel they "have to carry out the job twice". This is explainable considering their typical design workflow (adding asset after asset after the progressive evaluation of the concept interim states, i.e. along its evolution). For the preparation of a SPARK platform, they have adapted their standard workflow by adding an additional step, which is about the "decomposition" of concepts to the set of elementary assets that they are composed of, for which they also create variants to facilitate product evaluation and the application of desired modifications.
The development of the SPARK business model an exploitation plan should take this insight into account. For instance, it may be necessary to offer a service of training on how to generate variants and elementary assets in a more efficient way. This will be of particular importance for the packaging design market.

The designers at Stimulo felt there was no major change required in their design process. They noted that, when working on product interface design, the designers know in advance what kind of components are needed to visualize and interact with the product as they reflect the product functionalities that have been defined earlier in the process. Therefore, they can start creating basic assets in several variants without the need of generating a satisfactory overall composition for the interface and "disassembling" it in elementary pieces. Their design process of interface ideation and development mostly focuses on the placement of these visualization and interaction components to facilitate the user experience.

"*The SPARK GUI has now become easier to interact with*".

Designers at Artefice who had experienced the GUI during WP4 noticed an improvement in the functionalities the platform now provides and with the overall experience of using it to modify the mixed prototype, that has become more effective and versatile (see WP3 deliverables). Despite requiring a little time to adjust to the changes, after 30 minutes of interaction, the learning curve is almost complete and the designers feel confident in its usage.
Improvements are however expected in terms of a more effective integration (e.g. file exchange) with software commonly used to develop assets for packaging design (Adobe Illustrator, Photoshop…) as well as for what concerns a more precise management of assets onto the prototype to make it more precise (i.e. "stick to grid" option, fine-tuning of parameters…)
As none of the interviewees from Stimulo directly interacted with the GUI, they did not spontaneously mention anything specific about the capabilities of the GUI and the SPARK platform functionalities.

"The preparation of the "UV map" is something that we are not able of doing by ourselves".

This specific aspect is shared both in Artefice and Stimulo. Respondents 4 and 5 had the chance to prepare the UV map for the SPARK co-creative design session (one during WP4 experiments, one for WP5 ones) and they both explicitly mentioned difficulties for this specific activity. Respondent number 5 had to prepare the UV map for two sessions, while 4 did it just once. Respondent 5, when asked about the learning curve of this process, stated that the first time it is very difficult, then it gets better as one gets familiar with the procedure and it also takes less time. Still, there are some difficulties in surface reconstruction, particularly for complex featured surfaces. Also, this aspect should be seriously considered from the perspective of the business model for the commercial exploitation of the SPARK platform.

In both companies, interviewees declared that the SPARK platform introduces positive aspects within their design process. On the one hand, it allows a more effective interaction between designers and clients, as the mixed prototype enables a more effective communication and unleash opportunities for simulation of interaction with the product/prototype.

On the other hand, designers also perceived that the SPARK platform allowed a reduction of the whole development process duration, due to the lower amount of de-briefing sessions needed before converging towards a consolidated packaging concept. The savings for the Pizza and the Parmigiano sessions can be estimated in 2 de-briefing meeting less (which is approximately one working week + the waiting time for receiving the client's answer).
The Mozzarella session was at a more advanced stage of the project when the experimental session was conducted, but the SPARK platform allowed a real size version of the packaging to be visualized by designers and clients. It clearly showed that some assets should have a different size, as they were represented smaller than expected, somehow preventing a potential cause of inefficient design iteration.

4.3 Results of the session performance metrics

The results of the application of the session performance metrics are presented in Table 7 (Artefice) and Table 8. (Stimulo), below.
In both tables, the results from the sessions completed within T5.1 are shown on the left half of the table whilst the results from the three sessions completed within T4.2 are shown on the right of the table (in grey) for reference - note that only the 'SAR' session from WP4 is directly comparable with the WP5 sessions as the other WP4 sessions used different technologies for the creation of design representations.

Table 7: Results of the session performance metrics for the Artefice sessions.

Metric title	**Artefice – WP5**			**Artefice – WP4**		
	Pizza	**Mozzarella**	**Parmigiano**	**SAR**	**AR**	**No-ICT**
Duration	*87 mins*	*36 mins*	*44 mins*	-	-	-
Session objectives	*Filtering 9* *Generating 5*	*Filtering 9* *Generating 1*	*Filtering 9* *Generating 3*	*Filtering 10* *Generating 8*	*Filtering 8* *Generating 10*	*Filtering 8* *Generating 2*
Quantity of ideas	4	3	**5**	11	4	5
Variety of ideas	Coverage = 3 New = 0	Coverage = **3** New = **1**	Coverage = 1 New = 1	Coverage = 2 New = 0	Coverage = 4 New = 0	Coverage = 5 New = 1
Quality of ideas	**4**	2	2	3	1	2
Novelty of ideas	=25/4= **6.3**	=12/3= 4	=13/5= 2.6	=7/3= 2.3	=9/4= 2.3	=19/5= 3.8
Task Progress	1xHigh =3 4xMed = 8 2xLow = 2 Total= **13**	2xMed = 4 Total = 4	2xHigh = 6 Total = 6	2xHigh =6 Total=6	1xHigh =3 Total = 3	1xHigh = 3 1xMed = 2 1xLow = 1 Total = 6

Filtering Effectiveness	= 0/(4-2) = 0	=1/(3-1) = 0.5	=3/(5-2) = 1	=8/(11-1)=0.8	=3/(4-2)=1.5	=3/(5-3)= 1.5
Usability	54/100	23/100 11/100 Avg 17/100	46/100 41/100 Avg 44/100	70/100	59/100	32/100 71/100 Avg 52/100

The first two rows of Table 7 are not metrics but do help to characterise the sessions to aid comparisons. The first row shows the duration of each of the sessions. The second row shows the 'session objectives'. These scores came from the pre-session interview with the session facilitator who rated the importance of filtering and generating ideas on a scale of 1 to 10, where 10 is critical importance and 1 is very low importance. We can see that the three WP5 sessions all placed great importance on idea filtering (9/10), whilst low to moderate importance was placed on idea generation (1-5/10).

For the Quantity metric, the Parmigiano session scored highest with five new ideas generated. However, this is significantly less than the SAR session completed in WP4 in which 11 ideas were generated. However, it should also be noted that the WP4 SAR session placed greater importance on idea generation (8/10) than any of the WP5 sessions - which might explain this difference.

Concerning the Variety metric, it was the mozzarella session that performed best, with ideas appearing within three pre-existing rows of the morphological chart and one new feature row. For the Quality metric, the Pizza session performed best, with four ideas taken forward for further development. Those ideas were also highly rated in terms of novelty, with an average score of 6.3 for the pizza session.

The pizza session also scored very well in terms of the Task Progress metric, with a score of 13. It should be noted that the pizza session was more than double the duration of the other two sessions and therefore it is to be expected that more progress on project tasks would be achieved.

Concerning Filtering Effectiveness, it was the Parmigiano session that scored highest with a score of one, having filtered out three ideas to end up with the desired two ideas to take forward.

Finally, concerning usability, it was the Pizza the session that scored highest with 54, whilst the Mozzarella session scored very poorly, with an average score of 17. This very low CSI score can be explained by a number of factors:

- The concepts being tested featured images of the mozzarella on a white background, which were difficult to see due to the brightness of the SAR rendering.
- The mozzarella packaging was relatively small, which highlighted the limitations of the SAR rendering in terms of text legibility and pixelation of small graphics.
- The project had already advanced further than the other projects and so there was less interest in generating new ideas, which is confirmed by the score given for the importance of idea generation (one out of ten).
- It was the last session of the day and so the motivation of the participants might have begun to decline at this point.

Overall, it would appear that the Pizza session was the most effective in terms of generating high quality, novel ideas and making progress on project tasks. The success of this session might be related to the geometry and size of the prototype as it featured a large flat surface that resulted in very good rendering quality by the SAR technology.

Table 8: Results of the session performance metrics for the Stimulo sessions.

Metric title	WP5		WP4		
Session	**LISN**	**Neosonics**	**SAR**	**AR**	**No-ICT**
Duration	*45 mins*	*34 mins*	-	-	-
Session objectives	*Filtering 5* *Generating 3*	*Filtering 5* *Generating 8*	Filtering 7 Generating 7	*Filtering 5* *Generating 8*	*Filtering 8* *Generating 1*
Quantity of ideas	**8**	7	8	8	6
Variety of ideas	Coverage =**8** New = **2**	Coverage = 5 New = 0	Coverage = 5 New = 1	Coverage = 1 New = 1	Coverage=4 New=1
Quality of ideas	4	**7**	4	5	1
Novelty of ideas	=39/8=4.9	=34/7=4.9	=44/8=5.5	=51/8=6.4	=23/6=3.8
Task Progress	0	0	1xHigh = 3 Total = 3	2xHigh = 6 1xMed = 2 Total = 8	1xMed = 2 Total = 2
Filtering Effectiveness	= 4/(8-1) = **0.6**	0	= 4/(8-1) = 0.6	=3/(8-5)=1	=5/(6-1)=1
Usability	**84**/100	72/100	64/100	90/100	44/100

For the Quantity metric, the LISN session performed slightly better with eight ideas compared to seven in the Neosonics session. This is somewhat surprising given that in the session objectives there was a greater importance placed on idea generation for the Neosonics session (importance of eight versus three for LISN).

The LISN session also performed very well against the Variety metric, in fact the scores of Variety (Coverage) 8 and Variety (New rows) 2 were significantly better than any previous session completed by Stimulo within the SPARK project.

For the Quality metric, the Neosonics session scored the best as all seven ideas generated during that session will be taken forward for further consideration and development within the project. However, the fact that none of the ideas was filtered out does mean that the Filtering Effectiveness score for the session was zero. It was, therefore, the LISN session that performed best with a Filtering Effectiveness score of 0.6.

Both sessions scored the same on Novelty (4.9) and Task Progress (0) metrics. Whilst the Novelty score was somewhat lower than the SAR session completed during WP4 (5.5), the Task Progress scores were significantly worse. It is not clear at this stage why the Task Progress scores were so low.

For Stimulo, the best session overall appears to be the LISN session, which scored best on the Quantity, Variety, and Filtering Effectiveness. This is reflected in the high user satisfaction score provided for the Usability metric (84/100).

4.4 Results of artefact based interaction analysis

This section focuses on the results from the interaction based on artefacts analysis. The quantitative analysis was based on the results obtained with the on-the-fly coding method applied to five co-creative sessions presented before.

4.4.1 Co-creative sessions: Artefice case studies

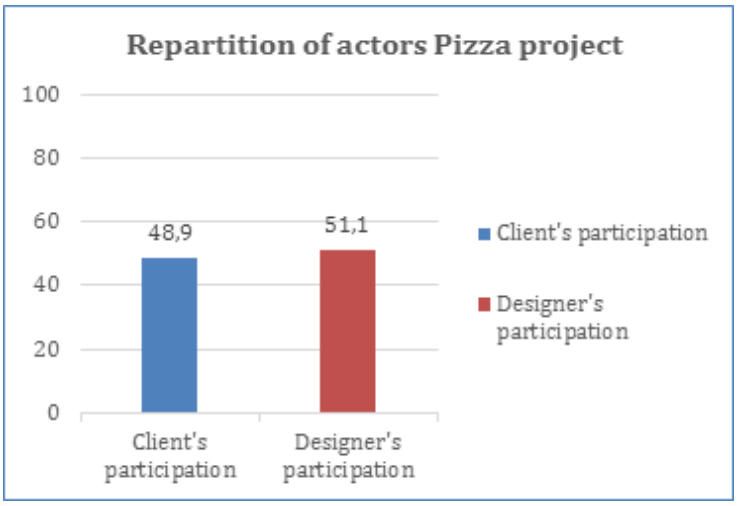
We start the analysis with the sessions recorded in Artefice. The three different sessions have been introduced in section 3.2. They gathered designers and clients, which is different from the other session observed in WP1 and WP4 where the sessions were involving designers and end-users.
In Table 9, we list the different artefacts involved for each session.

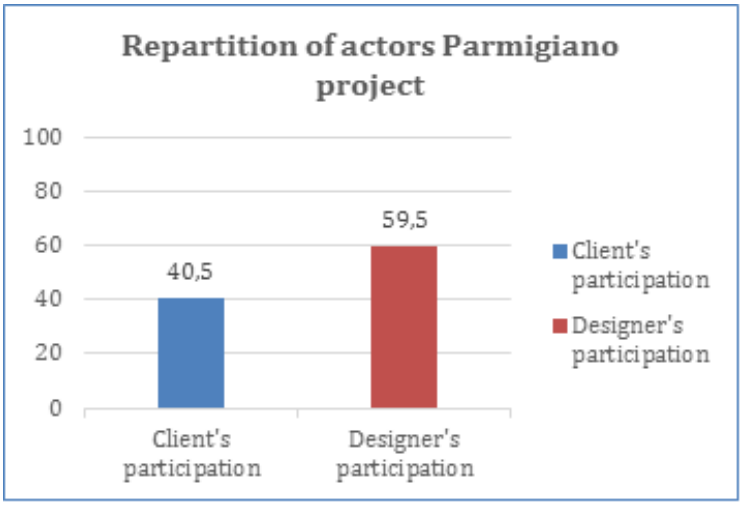
	Artefacts used	Screenshot from the session
Pizza session	SAR system: touchscreen as UI + mixed prototype Tangible: design representation of pizza proposals, bloc notes	
Parmigiano session	SAR system: touchscreen as UI + mixed prototype Tangible: design representation, bloc notes, post-it Digital: smart phone	

Mozzarella Session	SAR system: touchscreen as UI + mixed prototype Tangible: design representation of pizza proposals, bloc notes, printed 3D objects. Digital: smart phone	

Table 9: list of the different artefacts involved for Artefice session

Here we aim to compare the percentages of interactions initiated by Designers and Clients in order to evaluate and compare their participation. We observe that designers perform a higher percentage of the interactions in the three sessions (51-60%).
However, clients initiate a significant proportion of interactions in all the sessions (between 40% and 49%). This tends to demonstrate that the sessions are real co-design sessions and that the SAR stimulates the involvement of clients in the design task.





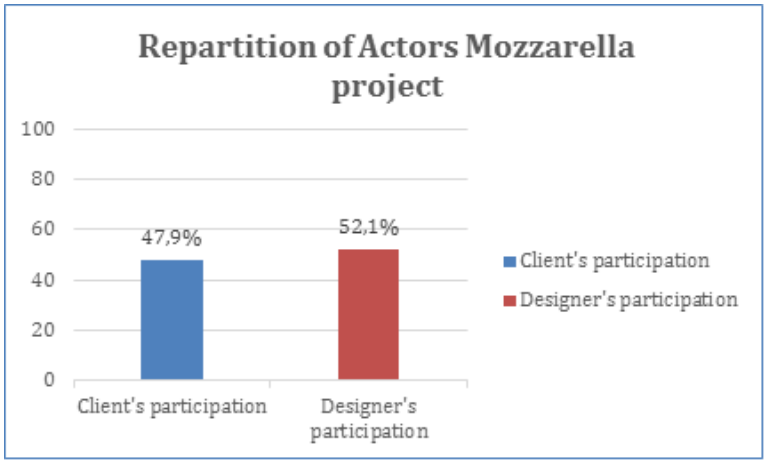


Figure 06. Repartition of Actors' participation in the different Artefice sessions.

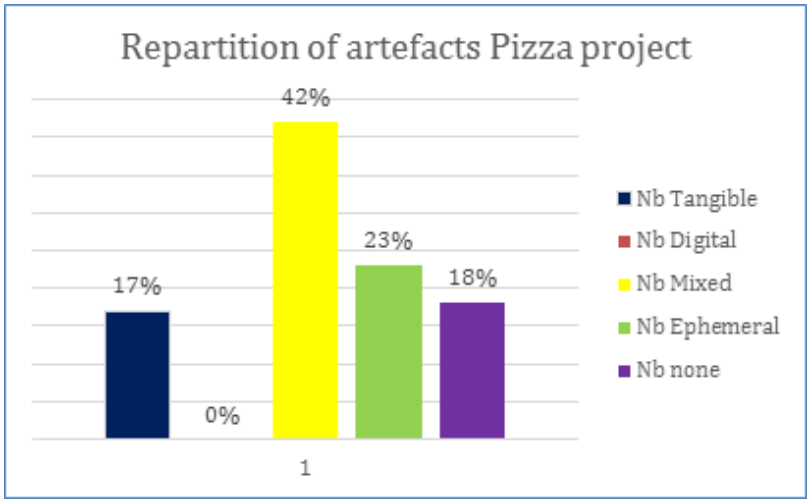
We consider now the types of artefact involved during the design sessions in order to evaluate their respective importance. Then we computed the proportions of each interaction type in each session. We note a very important percentage of mixed artefacts for all sessions (30 to 44%). The focus was on the SAR prototype for the whole duration of the session. The exchange between different actors was mainly animated with mixed artefact despite the presence of other types of artefacts in some moment of the meeting. This is an encouraging result that demonstrates the salience of the mixed artefact during these sessions.

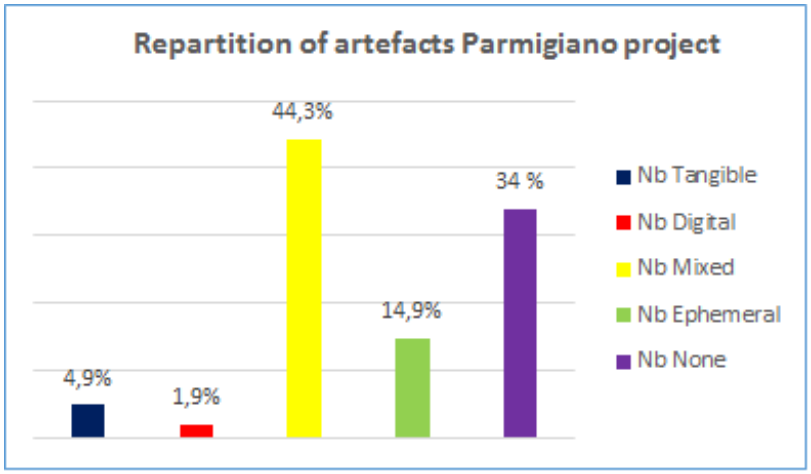
We also notice an important use of tangible artefacts, especially in the Mozzarella case. This corresponds to the introduction of printed design representations that were used to display other concepts than the one presented on the mixed artefact. This produced a concurrent centre of attraction and naturally the interactions shifted from one to the other media. This accounted for up to 40% of the total interactions. Further analysis of the verbalisation and task assignment could allow to understand more this shift and explain why the clients had to consider more the tangible artefact than the mixed one.

The "none" category also accounts for an important percentage of the total interactions in these results. This corresponds to phone calls or other interruptions that were not mainly related to the object of the session. In some cases, the presence of none artefact is due to technical problems such as the crash of SAR platform which implies the intervention of technical support to fix the bug and re-launch the session.

Gesturing in the air is still very present and, as in the other experiments, proved to be a good way of expressing ideas and support actively the discourse. It represents 15% to 23% of total occurrences during the session.
We can also discuss a kind of proportional use of gesturing during the session and the use of mixed prototype. Parmigiano session presents the highest level of mixed prototype involvement and the lowest rate of ephemeral artefacts which is the opposite case of Mozzarella project which featured the lowest use of the SAR artefact across the three sessions and showed an important rate of ephemeral gestures. This can lead to the hypothesis of the necessity to introduce the SAR system in order to reduce the ephemeral gestures to allow sessions' actors to better explain their needs and explicitly discuss the proposal.

We notice a very low percentage of interactions with digital artefacts sessions (0% to 1,9%) although the user interface of the SAR was significant (multitouch screen). We suppose that participants didn't feel the need to use digital means in the presence of the SAR mixed prototype.





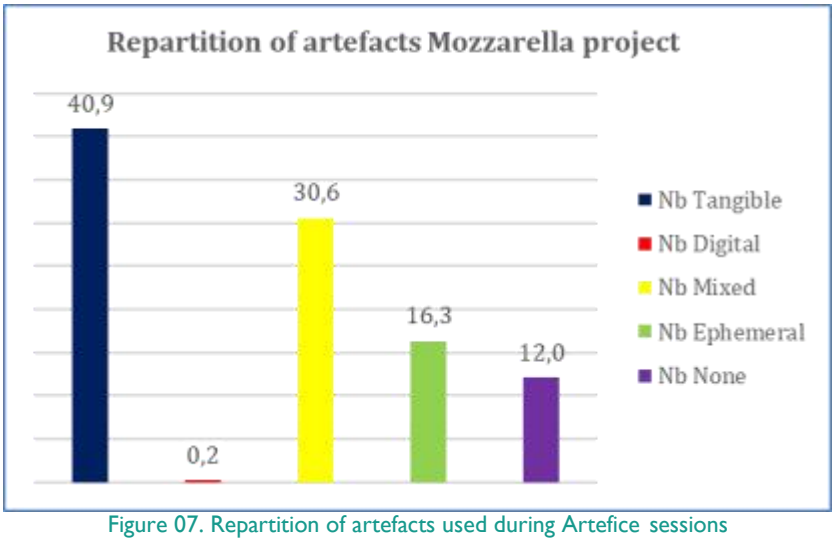


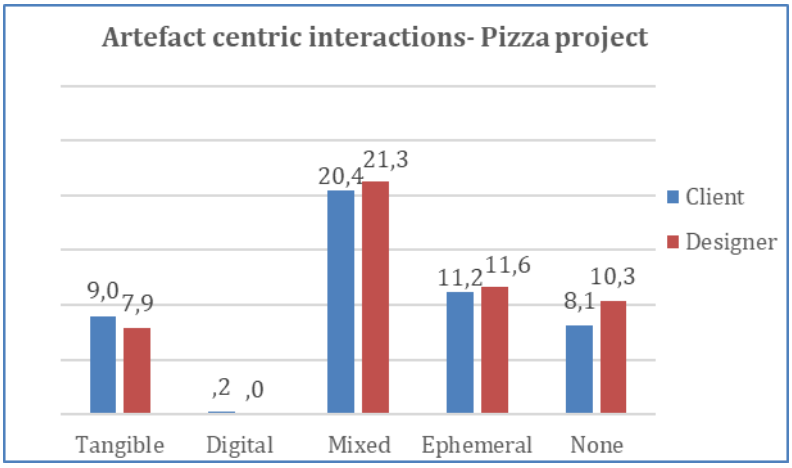
Figure 07. Repartition of artefacts used during Artefice sessions

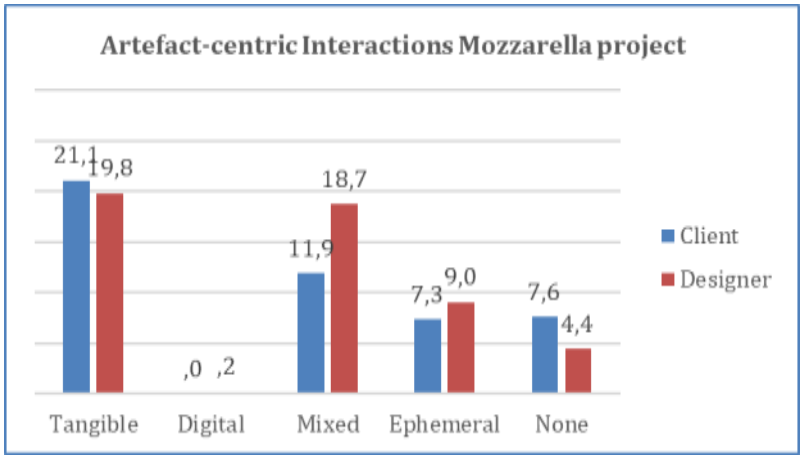
We compare now the respective use of artefacts by clients and designers. For that, we computed the proportions of the types of artefact used by each participant type in the different sessions. The main objective was to compare the respective proportion or interactions between the clients and the designers.

Since the mixed prototype is the main artefact proposed at the beginning of the session, it presents an important proportion of use, mostly employed from the designers who basically manage the session and manipulate the SAR interface. But we still notice a very good involvement of the clients, particularly in the Pizza project. Again, here an analysis of the verbalisations could help to understand why, in that case, the clients behaved slightly differently than in the other sessions.

The proportion of tangible artefact used is approximately the same for clients and designers. We can notice that the introduction of tangible artefacts has reduced the focus on the mixed prototype for both actors as shown on Pizza and Mozzarella projects, which is not the case of Parmigiano project characterised by a high average of mixed artefact use. For the three project, the tangible artefacts were always more used by clients, this may be explained by the fact of accessibility of these elements closer to the client and also by the fact that the designer was mainly manipulating the SAR interface.

An approximately equal rate is observed for ephemeral artefacts for the two categories of participants. This is also in line with the first results that show an even repartition of interactions between the two categories. Ephemeral artefacts seem to follow the same tendency than the total interactions.





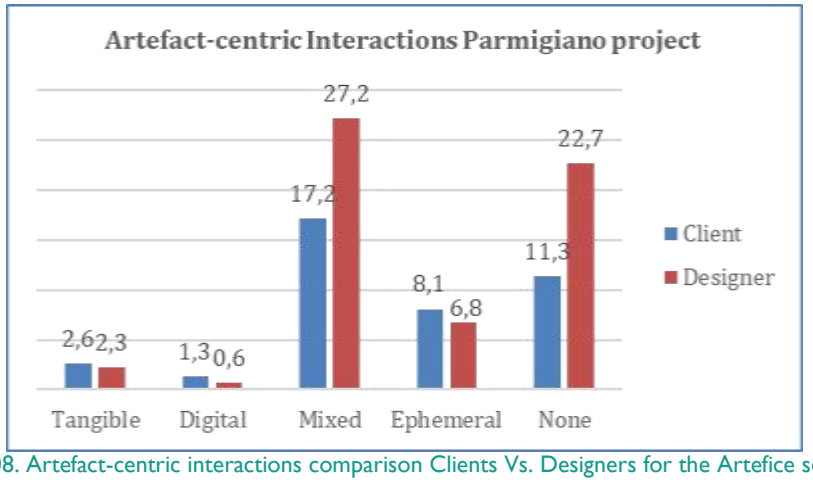


Figure 08. Artefact-centric interactions comparison Clients Vs. Designers for the Artefice sessions.

4.4.2 Co-creative sessions: Stimulo case studies

We will now look at the results of the observations made in the Stimulo agency. The two projects were in an early stage of the design process. Since Stimulo has decided to switch to a more agile way of work, they involve the SAR system in their task process. It aims at facilitating the exchange with their clients and give an opportunity to manage more product iterations.
First, let's mention which type of artefact was used for both sessions:

	Artefacts used	Screenshot from the session
LISN	SAR system: mouse + screen as UI + mixed prototype Tangible: bloc notes Digital: Smart phone	
Neosonic	SAR system: mouse + screen as UI + mixed prototype Tangible: a doll, bloc notes Digital: Smart phone	

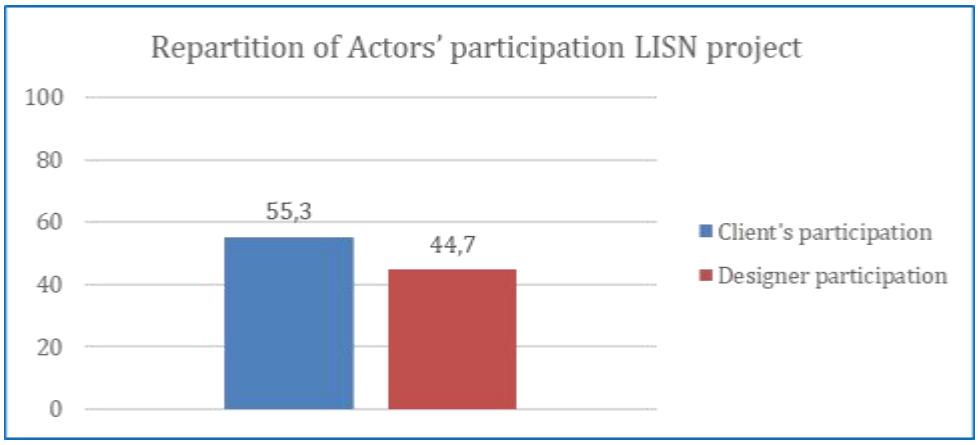
Table 10: list of the different artefacts involved in Stimulo sessions

Participation of clients in the co-design sessions

As for Artefice, let's consider the global participation of the clients and designers in the sessions. The diagrams below display the percentage of each actor participation during Stimulo sessions. Clients clearly dominate the session discussions with a higher rate going from 55.3% up to 69.9% of the interactions.

The contribution of designers is still reasonable (30.1% to 44.7%). Their main task is not to lead the session but to animate the discussion and manipulate the SAR interface to reflect the suggestions made by the clients.

We can identify an important implication of the clients during discussion This is the opposite trend to Artefice. Without more data, it is difficult to conclude more than that the effect of the context of the design sessions is important in the involvement of the parties. As we explained in the description of the sessions (section 3.2), the aim of Stimulo was to create lots of design scenarios, online, and therefore it created an important dynamic of change on the prototype which probably stimulated the clients.



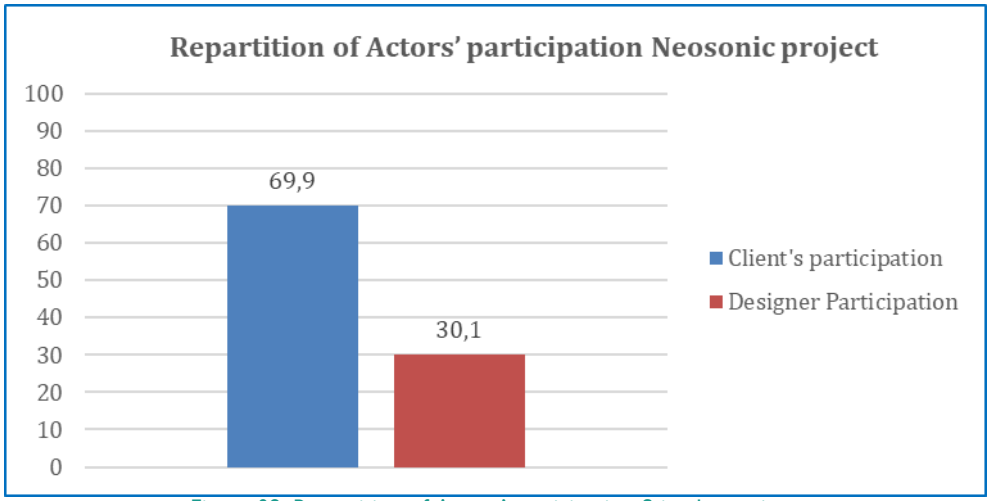


Figure 09. Repartition of Actors' participation Stimulo sessions

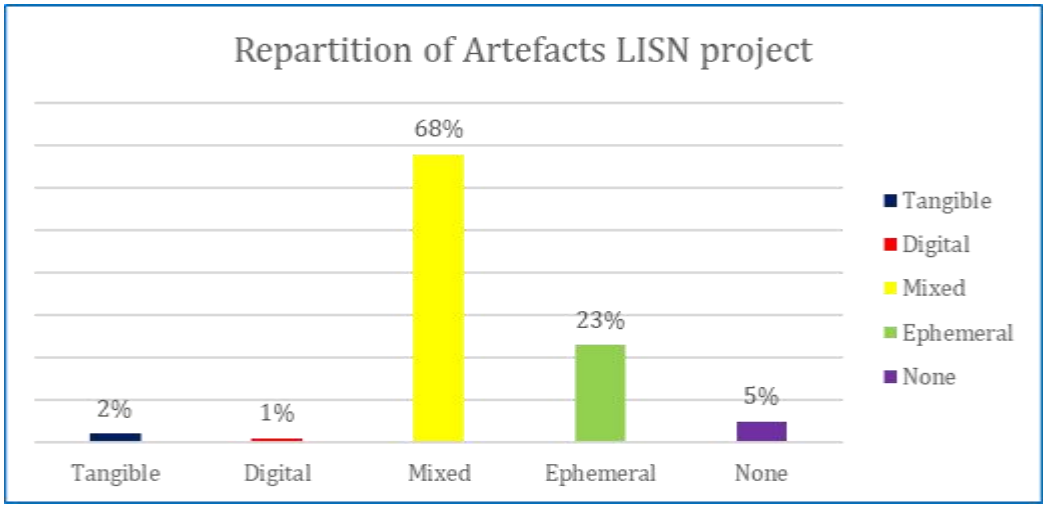
Regarding the repartition of artefacts involved during the design meeting, we computed the proportions of each interaction type in each session.

As shown on the diagrams, a very important percentage of mixed prototype is present for both sessions (57% to 68%) due to the focus on the SAR projected prototype which is the principal object of discussion between participants. Tangible interactions are mainly related to notes taking and sketching by the designer for Neosonic case.

The domination of mixed prototype against digital and tangible artefacts 1% to 2% for digital and 2% to 7% for tangible clearly shows a clear preference for this media against digital artefacts as they were also available through screen visualisation but not used. Again, further analysis is necessary to explain the reasons for this preference, but the hypothesis that was at the origin of the project seems comforted here. This can lead to the hypothesis that participants didn't feel the need to point at or discuss on any digital means giving the presence to SAR mixed prototype, or possibly they found the mixed prototype as enabler of a richer interaction with the design concept under discussion. Only some designers' notes were taken to gather the clients' feedback for Neosonic case.

Ephemeral artefacts were used though gesturing allowing participants to express ideas and needs with about 23% to 26% of total session occurrences.

For both cases, a low rate of none artefact (5% to 9%) is due to the global focus on using SAR system and absence of intrusive interruption of the running of the meeting.



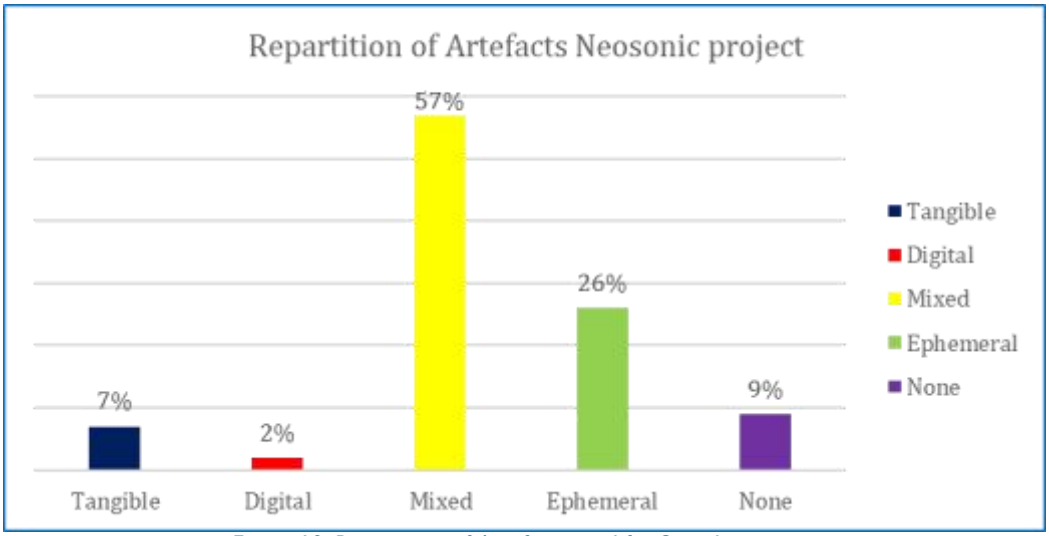


Figure 10. Repartition of Artefacts used for Stimulo sessions

The main aim here is to show which kind of artefact is mostly used by which type of participant (designer or client) during the session. For that reason, we computed the proportions of each interaction type for each participant in the same way as Artefice.

As shown in the figure 09, the mixed prototype is the most used type of artefact by both actors. In Stimulo cases, mixed artefact is mostly employed by clients. Clients showed an interesting preference in the use of the mixed artefact and intensively referred to this media during the meetings. This is a very encouraging result as it tends to demonstrate the superiority of the SAR as an interaction media against digital media. Additionally, Clients were capable of manipulating the prototype all along the session. Apparently, the mixed artefact was intensively used for testing scenarios and configurations that were projected in real time, and remained the focal point all along the session.

The proportion of tangible artefacts used is approximatively the same for clients and designers for the LISN session and is very low (1.2% and 1.1%). Tangible artefacts were used more by the Client in the Neosonic session (6% vs <1%), but overall remains low. We can notice that the digital artefacts were almost completely absent in both session. None of the actors felt the need of digital means to express their ideas.

The use of ephemeral artefacts shows similar trends across both sessions, with the Client's use being slightly higher than the Designers in both cases. This is a general trend we see all along our experiment for Stimulo and Artefice. This surprising result needs more investigation to understand the reasons of this stability.

The Neosonic session showed the highest percentage of mixed artefact use 42% for clients present also the highest rate of ephemeral artefact employed by both clients 15% and designers 10%. The LISN session present the same proportion of mixed-ephemeral artefact use.

For Clients the proposition of ephemeral artefacts is significantly low compared to mixed ones. For Neosonic session the designer interacted significantly less giving the priority to the client. The designer was taking notes in his notebook while the client was manipulating the mixed artefact.

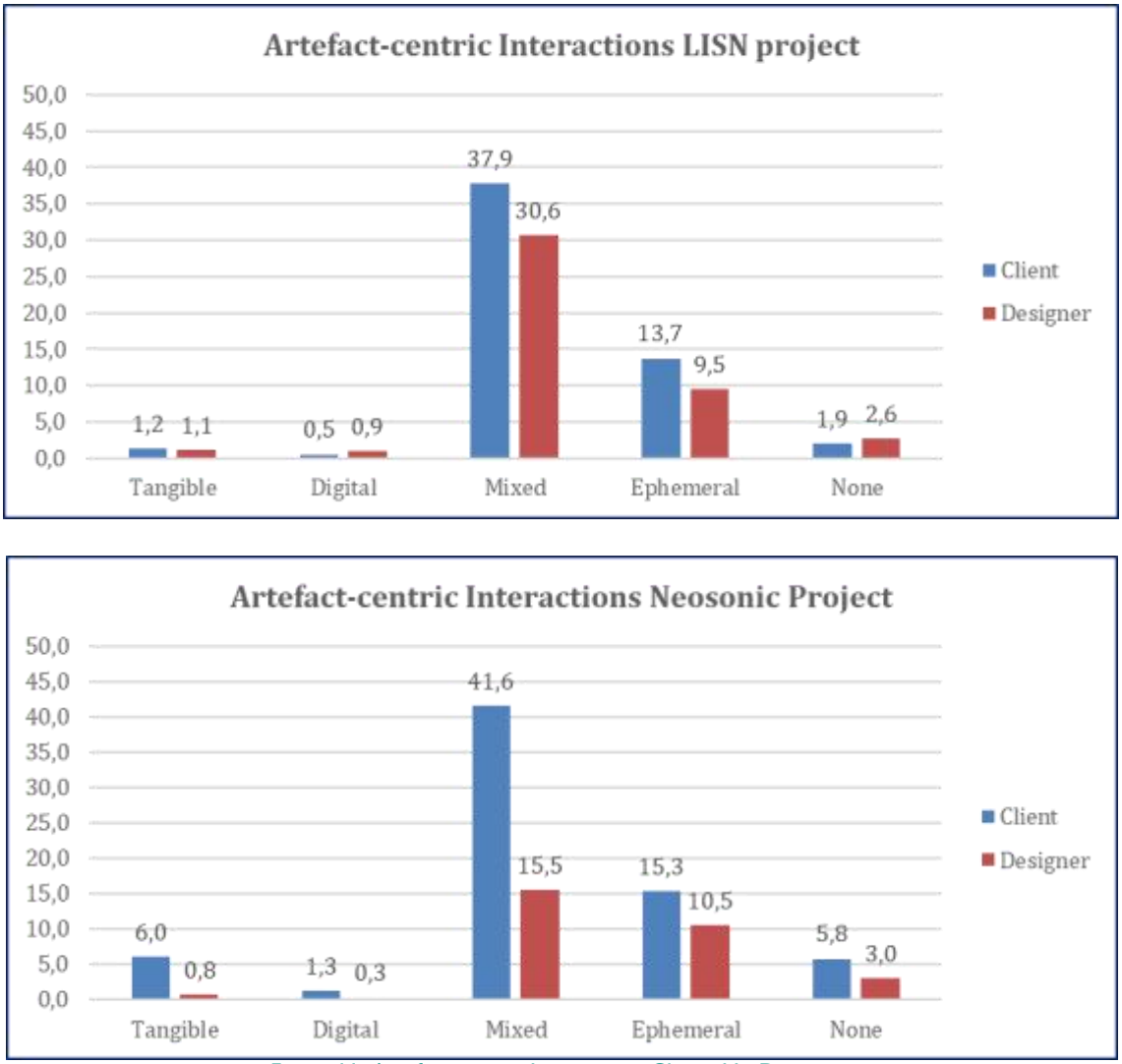


Figure 11. Artefact-centric Interactions Clients Vs. Designers

4.5 Results of the logfile analysis

Figures 12 and 13 respectively present the application of the metrics presented in Section 3.3 for the analysis of the log file captured during the sessions as the co-designers interacted with the SPARK platform. These tables, collect both the data of WP5 and relevant WP4 experiments. This is to help making comparisons within WP5 sessions and between WPs.

The last column of the tables, then, reports figures for what observed during the sessions supported by SPARK (SAR testing conditions) along WP4. The data therein presented differ from each other in that the available data for Artefice's WP4 session come from an analysis of the log file, while Stimulo's refers to the outcome of standard design protocol analysis (as the logfile recorded during that session was corrupted).

ARTEFICE	Pizza	Parmigiano	Mozzarella	SAR ARTEFICE T4.3/4.4 (LOG ANALYSIS)
E1	63,3%	63,3%	60,0%	46,7%
E2	76,0%	63,3%	100,0%	57,6%
E3	120,118	75,072	175,655	82,816
Q	285	65	63	181
V	0,067	0,292	0,238	0,271
Duration	2822	1426	2635	4058

Figure 12: results for the metrics describing the design process after the analysis of logfiles - Artefice

STIMULO	Neosonic	LISN	SAR STIMULO T4.3/4.4 (DESIGN PROTOCOL ANALYSIS)
E1	20%	20%	8%
E2	33%	33%	21,2%
E3	409,324	539,7	454
Q	47	106	52
V	0,106	0,0471	0,026
Duration	2047	2699	3178

Figure 13: results for the metrics describing the design process after the analysis of logfiles - Stimulo

Duration of the session

T5.1 sessions have variable durations both within and between the domains (packaging and product interface design). They also present variable durations compared to what recorded in WP4.

For the Artefice sessions, the most appropriate comparison with the WP4 sessions (in terms of duration) should be Pizza vs SAR WP4 (Pumpkin soup). In both these sessions. the co-designers explored three different pre-conceived alternatives and co-worked to re-elaborate them and create completely or partially new concepts. The figures about the duration suggests that probably the UI has introduced significant improvements in terms of usability as the session lasted approximately 20' less, which is approximately -30%. On the other hand, it is also possible to notice that WP5 sessions involved designers that already took part in the WP4 sessions. This might show some effects due to progresses in learning when interacting with the platform and its UI (learning curve progress).

For what concerns the Pizza session (Figure 12), it is worth mentioning that the duration refers to the second part of the session with clients, as the first part was almost exclusively focused on the evaluation of pre-conceived design proposals (almost pure review). The second part of the pizza session here considered required co-designers to actively participate in the definition of new proposals and therefore appears to be the right candidate for log file analysis. The Pizza session log file was initially divided into two halves, consistently with the nature of the session, as the SPARK platform was rebooted between the review and the co-creative part of the activities.

The other figures concerning the durations of the sessions substantially confirm what emerged for the Pizza session, as the duration of the session is somehow of the same order of magnitude and significantly lower (-50% wrt pizza) for the Parmigiano session.
Stimulo's sessions also confirms the above mentioned trend. Despite the order of magnitude remained the same, the case study of the Wave control system (data in last column of Figure 13, see D4.2 for a complete description of the case study) had a longer duration. The experimental sessions of T5.1 lasted on average the 25% less of what recorded in WP4, using the SPARK platform. A more precise comparison could be done between WP4 and Neosonic data, as both the sessions focused on a handheld product that provides information to the user. In such a case, the time savings are approximately 35% (the Stimulo designer that interacted with the GUI during the co-creative sessions of T5.1 also participated in WP4 experiments with the same role).

Exploration of the design space
The index E1 provides a measure of the degree of utilization of the SPARK platform with reference to the set of item-parameter couples the co-designers can interact with during the sessions. The denominator of the ratio (see Sect 3.3) counts all the possible combinations the platform allows to deal with in terms of functionalities it offers to the user. WP5 experiments show that there is a significant increase of this degree of utilization of the platform potential. Across the three observations at Artefice, the results are significantly stable and higher with reference to what was recorded during WP4 (45% in WP4 to 60% in WP5 is an improvement of approx. 33%).
For what concerns the Stimulo sessions, the trend is quite similar despite the ratio is significantly lower between the two domains (lower for product interface, higher for packaging design). In fact, for both the session the index E1 scores 20%. This value, however, is +150% of what recorded in WP4.
These two results provided consistent outcomes for the index E1. The consortium wonders whether this can depend on the profile of the participants in the co-design session. As explained in Section 3.2, the profile of designers is similar between WP4 and WP5 and some of the designers involved in T5.1 also took part in WP4 experiments. Nevertheless, the different profile and personality of clients might be an uncontrolled (and uncontrollable) biasing factor.
The discussed improvements for the duration of sessions (under the assumption that shorter durations correspond to time savings for the design process as a whole) might be also due to the changes and additional functionalities that the WP5 version of the SPARK platform UI presented, as that release already addressed some of the most pressing requests for improved usability, as emerged after WP4 experiments and interviews with designers.

E2 differs from the previous index in that it measures the degree of utilization of the platform with reference to the actual set of possible item types prepared by the designers for the session, which is a smaller set than the whole set of items the platform can handle (i.e a subset). As for E1, it is expressed as a ratio. In other words, E2 describes the degree of exploration of the SPARK platform according to the classes of items for the available assets.
For what concerns the Artefice sessions, the WP4-WP5 comparison shows that there is a general improvement in the exploration of the design space with the assets available in the platform. This increment, however is largely variable along the different sessions. This could be a consequence of the nature of the considered case studies (results are more similar b/w Pizza and Parmigiano, less for Mozzarella.) This is also consistent with what emerged during the interviews with designers, which declared they had to prepare larger set of items for projects that are at earlier stages of development.
For what concerns the Stimulo sessions, the WP4-WP5 comparison is the only one meaningful (WP data are the same between the two T5.1 sessions). The increment for E2 is not as marked as for E1, but the

exploration recorded through the degree of utilization of the platform shows that this value shifted from 21,2% to 33%, which is an increment of approximately 60%.

E3 measures (indirectly) the persistence interacting with items, which was particularly meaningful to compare different testing conditions in WP4. It averages the time spent on each of the explored item-parameter couples
For what concerns packaging design (Artefice), the results show a very large variability among the WP5 sessions and also with reference to what was observed in WP4. The consortium assumes this index to be particularly context- (case study-) dependent and its meaningfulness for the analysis of WP5 becomes questionable.
In the co-creative sessions of product interface design (Stimulo), the results are definitely more homogeneous. The average between the WP5 sessions differs from what recorded in WP4 of 5% (while in Artefice's sessions the difference between WP4 and WP5 is approx. 50%). This result calls into question the previous statement on context-related biasing factors.

Quantity of Ideas
The Quantity of ideas (Q) is computed as the amount of changes carried out along the session with the assets managed by the SPARK platform. It accounts for everything recorded in the log that reflects an intention of changing the mixed prototype.
For the Artefice sessions, Q largely varies across the different case studies (both WP4 an WP5). The results scored by two sessions so different in nature (according to the progress of the different project Parmigiano and Mozzarella) produced similar results, which are much lower (approximately -65%) compared to the WP4 session. On the contrary, the Pizza session produced the most impressive results in terms of generation of new ideas. This phenomenon is probably due to:

- a larger productivity of the participants (first session, fresh mind and ideas),
- a larger canvas to work with, as smaller surfaces allow less changes that becomes visible and appraisable by sight, as for packaging (communication) design.

For the Stimulo sessions, the trend appears to be similar to the packaging design field. This can be clearly visible through the difference of Q values in WP5 experiments, which is of the same order of magnitude of what measured in WP4.

To reduce the bias of the variability across the considered sessions, a relative measure of quantity can be computed with reference to the duration of the sessions themselves (Q/duration).
For what concerns Artifice these results, show that the session of Parmigiano is more consistent with the WP4 session (0,045 vs 0,044).
The different nature of the Mozzarella session is now evident, as it produces a score that it is approximately half of the above reported (0,024).
The Pizza session still represents an outlier with reference to the Parmigiano and the WP4 session despite their similar nature (score 0,101 à +225%).
For the Stimulo sessions, the Q index normalized for duration of the sessions show results which are comparable with Artefice's (Pizza still remains an outlier). The comparison WP4-WP5 shows that the platform provided some improvements for idea generation: The WP4 session scored 0,016, while the T5.1 case studies scored higher values (0,023- Neosonic- and 0,039 -LISN).

The packaging for the Pumpkin Soup is a little larger than the Parmigiano one, but somehow comparable at least for what concerns the top surface. The Pizza packaging, on the contrary, communicate the food features just by means of the upper part of the pack, whose surface is approximately 400% more than the top surface of the pumpkin soup. The two devices (as mixed prototypes) used by Stimulo differed from

each other in size. The handheld device for health diagnostics is pretty similar in size to the Wave Control System of WP4. Nevertheless, the case Study by LISN involved a larger mixed prototype whose external surface represent a unique and large canvass for co-creation. The better results of Pizza and LISN respectively in Packaging and Product design for Q/Duration suggests that larger canvases to work with allows for a more effective idea generation.

Intensity of variations of applied modifications

The index V displays the intensity of the variations applied with reference to the overall number of applied modifications. As detailed in Section 3.3, smaller numbers for this index correspond to a more effective interaction with the platform and the assets therein managed.

Within the sessions carried out at Artefice, it is worth noting that the sessions concerning Parmigiano (WP5), Mozzarella (WP5) and pumpkin soup (WP4), as they all score close to 0,3. The results are very different for the Pizza session (WP5), where the record is 0,067 that is also a lower order of magnitude. For what concerns Stimulo, the design sessions provided a V which ranges between 0,016 in the WP4 session and 0,039 in the LISN case.

The sessions at Artefice and Stimulo provided the above ambiguous results for the index V. The ambiguity concerns the assumption that a larger canvas typically enables multiple subjects to interact and therefore enabling co-designers to interact more freely with assets, concepts and ideas. This still appears to be true in packaging design, but not particularly confirmed for product design interface, as the worst session in terms of score is the one with the largest mixed prototype.

4.6 Results of the design process efficiency metrics

At Artefice, the fresh pizza project was completed using the SPARK platform and the results evaluated using the design process efficiency metrics. At the time of writing, the fresh pizza project had reached the end of the 'ideas development' phase and was just waiting for one final decision from the client concerning the option to use the same graphical layout for both of the two pizza varieties or use the two separate layouts that had been developed. Given that the project had essentially completed the creative aspects, it was possible to proceed with the comparison with the benchmark data - which in this case was the frozen pizza project completed in 2017. A comparison of the design process efficiency metrics application results is presented in Table 10, below.

Table 11. Comparison of the design process efficiency results for the fresh pizza and frozen pizza projects.

Metric	Definition	Fresh Pizza (with SPARK)	Frozen Pizza	Change
Person-hours spent on project	All hours spent on project by design agency (including unbilled hours)	39	58.25	-33%
Lead time	Number of days between project start date and product launch date	21	100	-79%

Total development cost	Direct costs incurred by design agency	Confidential		-20%
Cost of prototype production	Cost of preparing all design representations used in collaborative sessions or sent to client (materials and labour)	Confidential		+47%
Re-work iterations	Total number of co-creative design sessions completed within the project	1	N/A	N/A
	Number of versions released up to end of creative phase.	4	8	-50%
	Average cost for each version release	Confidential		+34%

The results show that the Fresh Pizza project (completed with SPARK) made significant efficiency improvements across several metrics. The person-hours spent on the project between the initial briefing and the completion of the Ideas Development phase was reduced from 58.25 hours to 39 hours - a 33% reduction. This improvement was mainly due to the reduced number of iterations that were necessary (from eight down to four) to define and agree a final layout with the client. The reduced number of iterations also led to a major reduction in the project lead time (79% improvement) and a significant reduction in the total development cost (20% reduction). It should be noted that the cost of the SPARK installation has not been accounted for within the total development costs as, in practice, the SPARK installation would be used across many different projects and would therefore become an indirect cost for the company.

The cost of prototype production increased by 47%. This was mainly attributed to the 10.5 hours that were used to prepare the SPARK session (preparation of assets, UV map and physical model in white). The insights from the interviews with designers, presented in the following sub-section, help to explain why this session preparation activity took so long. Given that the designer who prepared the session had not done this task before it is reasonable to assume that this figure of 10.5 hours to prepare one session will reduce as the designers gain experience of using the SPARK Information System to prepare sessions. Further reductions in session preparation time should be possible if some of the ideas for improvements to the Information System identified from the interviews with the designers are implemented.

The only other metric for which the Fresh Pizza project (with SPARK) performed worse than the benchmark was the average cost of each version release. This was again mainly due to the extra time (and hence cost) required to prepare the SPARK session.

At Stimulo, one session has been completed using the SPARK platform as part of the Neosonics longitudinal case study. The Neosonics project was not completed at the time of writing and so it was not possible to apply the design process efficiency metrics at Stimulo. Once the Neosonics project is complete, the metrics will be applied and the results will be presented in D5.3.

4.7 Discussion

In the previous sections we have presented the results from a variety of research activities and analysis methods. Here, we present a synthesis of some of the key points that can be drawn from across the results.

Preparation of the SAR sessions is challenging

From the interviews with designers, one of the key challenges during the preparation of the sessions was the generation of the 'UV map', which involves a fairly long and complex procedure and requires some technical understanding of augmented reality technology. Despite attempts to provide a step-by-step guide to this process, new users still required some support to complete the process. Considering the implications for the commercialization of the SPARK platform, it will be necessary to make this process significantly simpler or provide some type of enhanced technical support service (e.g. video guides, telephone technical support etc.).

Another challenge identified at Artefice was the time required to prepare the assets for use in a SPARK session. They felt that this was a significant additional step that would need to be integrated into their process. Digging deeper, it was discovered that the Artefice designers were creating a number of complete design proposals, and then decomposing those concepts into the constituent assets for use in the session. This suggests that the sessions completed at Artefice are more like 'creative review' sessions – which were a type of co-creative session identified in WP1 in which the focus is on reviewing existing concepts and using the feedback from the client to iterate the concepts – see Figure 13. This is in reinforced by the 'session objectives' data captured during the pre-session interview, in which, for the Artefice sessions the importance of 'idea filtering' was 9/10 for all three sessions whilst the importance of 'idea generation' was never rated higher than 5/10. These scores suggest that the Artefice designers were always expecting a strong emphasis on idea filtering, with a limited amount of idea generation.

Figure 14. Types of creative session and design activity that could be supported by the SPARK platform.

For Stimulo, the designers reported that no significant changes were required to the design process in order to integrate the SPARK session (beyond a little extra time required for preparation of the UV map). This different experience compared to Artefice was explained by the fact that Stimulo projects tended to focus on the layout of user interface elements. They therefore required a much more limited range of assets (e.g. a few alternative buttons, lights, and speakers) and did not required the designer to develop complete proposals before the session. They could instead start the session with just the basic model and a small range of assets.

Improvements in the SPARK platform are delivering better results

The results from the log file analysis showed a significant improvement in the Exploration and Quantity metrics for both companies compared to the SAR sessions completed in WP4. They also showed that the duration of the sessions had reduced, which suggests that designers are now able to show what they want and make changes faster than previously. Similarly, the session performance metrics revealed improvements in the Variety, Novelty and Task Progress metrics at Artefice and the Variety, Quality and Usability metrics at Stimulo. We attribute these objectives improvements in session performance to the enhancements made to the SPARK platform in the second release, particularly the user interface. This claim is reinforced by positive statements received from designers about the user interface during and after the sessions. In the formal interviews with designers, both companies stated that SPARK has the potential to improve the outcomes of their projects, either by improving the communication or by reducing the number of iterations required to arrive at a commonly agreed final design. Whilst there are clearly still areas for improvement, we can claim that the improvements made in the second release of the SPARK platform have generated better results from the sessions and improved usability from the perspective of the designers.

The mixed prototype was the preferred artefact for supporting interactions

The results from the artefact-based interaction analysis revealed that the mixed prototype, generated using the SPARK platform, was used significantly in all of the sessions to support interactions between the Designer and the Client (in the range 31-68% of all interactions). Furthermore, it was the most commonly used artefact in four out of the five sessions, despite the presence of other conventional types of design representation, such as paper-based drawings. It is particularly encouraging to see that the Client was able to make good use of the mixed prototype (average of 17% of Client interactions at Artefice and 40% at Stimulo). This therefore suggests that the SPARK platform is an effective technology for supporting communication between designers and clients within the context of co-design sessions.

A further interesting insight from the artefact-based, interaction analysis was the fact that a significant number of 'ephemeral' artefacts were employed in the interactions across sessions (15-26% of all artefacts used). These ephemeral artefacts can include communication gestures and gestures used to explain an idea without the use of any physical artefact. It seems that participants still feel the need to use ephemeral artefacts to help explain their ideas. It had previously been hypothesized that the use of ephemeral artefacts would decline as the use of the mixed prototype increased (as the enhanced support for communication offered by the mixed prototype should reduce the need for ephemeral artefacts). There is some evidence to support this hypothesis from the Stimulo sessions but the evidence from the Artefice sessions is more mixed. This is therefore a topic that deserves further study.

SPARK delivers significant cost savings and efficiency improvements across the design process

The results from the longitudinal case study completed at Artefice were very encouraging, with significant improvements in person-hours (-33%), lead time (-79%), total development cost (-20%) and number of version releases (-50%). Whilst there was an increase in the cost of prototype production (+50%), this was attributed to the extra time required for the preparation of the SPARK sessions – which should reduce significantly as designers gain more experience of this process.

There is another longitudinal case study currently in progress at Stimulo, the results of which will be reported in D5.3. It is hoped that this case will show similar improvements in design process efficiency. These data, beyond their scientific value in demonstrating the effectiveness of the SPARK platform, will provide powerful and persuasive evidence that can be used in the marketing material during the exploitation of the SPARK platform.

5. CONCLUSIONS

Task 5.1 required the SPARK consortium to conduct validation tests at end user premises on real-life case studies with the aim of responding to SPARK objective 3:

> *"Study and analyse how and to what extent the SAR technology can stimulate and enhance design creativity through a comparison against pre-defined metrics in real operational design environments"*

Accordingly, five sessions were completed at the premises of Artefice and Stimulo (due to the nature of its business and clients, AMS projects have been considered more suitable for activities related to Task 5.3). The performance metrics and analysis methods developed in T4.2 were applied to each of the sessions leading to a number of key conclusions concerning the support offered by SAR technology for design creativity.

First, the overall pattern of results suggests that **SAR technology is beneficial in stimulating design creativity**. This conclusion must be qualified by stating that the supporting data set remains relatively small and the improvements shown do not span all performance metrics, but in general, the test sessions completed in T5.1 have shown a number of significant performance improvements in comparison to the results of WP4. Further evidence of the benefits of SAR technology in a more diverse range of design applications will be obtained in Task 5.3 and presented in D5.3.

Secondly, **use of SAR technology has been shown to provide significant improvements in design process efficiency.** Again, this conclusion must be qualified by acknowledging that it is based on one case study, but the improvements made were significant and it is hoped that the results from the second longitudinal case study will confirm this conclusion.

Thirdly, the **SAR technology was heavily used during the sessions and appears to support enhanced communication between designers and clients/end-users**. The results from the artefact-based gesture analysis confirmed that the SAR technology was used significantly in all sessions and was the most used artefact type in four out of the five sessions. Furthermore, there was significant use of the SAR technology ('mixed artefacts') by the clients in each of the sessions. These data, along with comments from the participants support the belief that the use of the SAR technology is contributing to the quality of the communication within the session and to the success of the outcomes.

Beyond these conclusions concerning the effectiveness of SAR technology for design creativity, a number of other practical conclusions can be drawn:

- Further action is required to facilitate the session preparation activities, including the generation of the UV map and the preparation of the assets. This might include further developments to simplify the UV map creation process or the creation of training services to reduce the learning curve for designers learning how to prepare sessions.
- SAR technology is less effective for very small prototypes due to the limitations of the rendering quality. One way to address this challenge would be use the tablet-based, AR set-up (used in the WP4 tests) when dealing with very small prototypes. This idea is supported by the very good results that were achieved using the AR set-up for the 'smart fitness product' in the WP4 experiments – see D4.2 Results of the experiments benchmarking the platform.
- Whilst significant progress has been made in the SPARK user interface, there are a number of new features that should be considered for development within T3.2. These include:
    - More effective integration (e.g. file exchange) with software commonly used to develop assets for packaging design (e.g. Adobe Illustrator, Photoshop etc.);

- More precise management of assets when placing them on the prototype (i.e. "stick to grid" option);
- The ability to fine tune key parameters – such as rotation and scale of assets.

These practical conclusions will be taken into account during the final development activities and within the exploitation planning process.

REFERENCES

Cherry, E. and Latulipe, C. (2014) 'Quantifying the Creativity Support of Digital Tools through the Creativity Support Index', *ACM Transactions on Computer-Human Interaction*, 21(4), pp. 1–25. doi: 10.1145/2617588.