EXPLORING THE PERFORMANCE OF AUGMENTED REALITY TECHNOLOGIES IN CO-CREATIVE SESSIONS: INITIAL RESULTS FROM CONTROLLED EXPERIMENTS

J. A. O'Hare, E. Dekoninck, L. Giunta, J.-F. Boujut and N. Becattini

Abstract

Co-creation can offer business benefits such as increased speed to market, increased product quality, and a reduced risk of market failure. However, co-creative design sessions can be challenging due to communication barriers between designers and non-designers that can result in misunderstandings and inhibit the efficient exchange of ideas. The potential for augmented reality-based design representations to overcome these challenges and support more effective co-creation sessions is explored through controlled experiments conducted with professional designers and end users.

Keywords: co-design, augmented reality (AR), collaborative design, technology development, co-creation

1. Introduction

There has been a general trend in engineering design over many years towards greater involvement of end users within the design process. At one time, end users were rarely consulted, but approaches such as 'user-centered design' and now 'co-creation' have transformed the role of the end user from the 'subject' of the design process to a 'partner' in the design process (Sanders and Stappers, 2014). Key business benefits claimed by companies practising co-creation include: increased speed to market; increased product quality; and a reduced risk of innovation efforts not meeting customer needs (Business Innovation Observatory, 2014). But with these benefits comes a number of new challenges.

A prime example is the communication challenges that can occur when designers and non-designers are trying to work together in a co-creative session. For most products, the end user is not a designer, and is therefore unlikely to have the same design vocabulary and experience of the design process. This may lead to misunderstandings and frustration when the exchange of ideas is inhibited by communication challenges (Stacey and Eckert, 2003). There is therefore a need for new methods and tools that can support effective communication and idea exchange within co-creative sessions. Design representations play a crucial role in any form of design communication (Bodker, 1998) and so developing new types of design representation that can better support communication between designers and non-designers is one obvious avenue to explore to address these communication challenges.

With this objective in mind, there has been some interest in Augmented Reality (AR) and Spatial Augmented Reality (SAR) technologies and their potential application within the design process. Both AR and SAR technologies offer the capability to enhance a physical prototype with a digital overlay - which is visualised through a screen or headset in the case of AR and projected directly onto the

prototype in the case of SAR. Furthermore, the information provided through the digital overlay can be modified on the fly. By providing supplementary information through the digital overlay and the potential for on-the-fly modification, it is hypothesised that AR/SAR technologies will be an effective type of design representation as they can provide many of the benefits of virtual/digital design representations in terms of richness and flexibility, whilst retaining many of the important benefits of physical design representations, such as their tangibility and accurate representation of scale.

This paper reports on work completed within the SPARK project, which is exploring the potential application of AR/SAR technologies to create design representations for use in co-creative design sessions. We define a 'co-creative design session' as: a pre-arranged session that involves designers and people not trained in design, working together in the design development process. The project is developing the 'SPARK Platform', which enables designers to create SAR-based design representations and modify them on-the-fly during a co-creative design session. Having developed a first prototype, testing of the SPARK Platform has begun with the aim of establishing what benefits there are, if any, of employing this type of technology within co-creative design sessions. Within this context, the objectives of this study are: to begin gathering data comparing the effectiveness of co-creative design sessions conducted with AR/SAR-based design representations versus conventional design representations (e.g. sketches, 3D renders, foam models etc.); and to identify specific strengths and weaknesses of AR/SAR-based design representations in comparison with conventional design representations. The first of these objectives is addressed through the following research questions:

RQ1 - "Do co-creative sessions involving AR/SAR-based design representations result in more effective co-creative design sessions in terms of idea generation, progress on design tasks, and filtering of ideas than those conducted with conventional design representations?"

RQ2 - "Do designers perceive AR/SAR-based design representations to be more effective for co-creative design sessions than conventional design representations?"

Whilst it was not expected that this small-scale study would be able to provide definitive answers to these questions, positive responses to both of these questions will ultimately be necessary if designers are to consider using AR/SAR-based design representations within their co-creative sessions.

The second objective of this study is addressed through the following research question:

RQ3 - "What are the specific strengths and weaknesses of AR-SAR-based design representations in comparison with conventional design representations?"

Answers to this question will help to inform the future development of the SPARK Platform and are also relevant to technology developers and researchers interested in supporting co-creative design.

To answer these research questions, a series of controlled experiments were conducted in which professional designers from the product design and packaging design sectors worked on real-life projects with end users using AR-based, SAR-based, or conventional design representations. The designers and the real-life projects came from two industrial partners of the SPARK project: Stimulo, a product design consultancy; and Artefice, a brand communication and packaging design agency.

The following sections provide a review of previous attempts to use AR/SAR technologies in the design process along with further background on the SPARK project and the SAR technology being developed.

1.1. Previous work on the use of AR and SAR technologies in design

There have been a number of studies that have investigated the potential for AR technologies to support group design activities (designers working with other designers). In an early study in this field, Billinghurst et al. (2003) looked at task performance and communication patterns between pairs of designers using various AR and conventional tools to collaborate on an urban planning task. They found that the teams took longer to complete the task using the AR technology than in face-to-face or projector display conditions. There were also significant differences in communication patterns, with, for example, more questions asked in the AR condition. They conclude that many of the challenges related to AR technology were due to the limited field of view afforded by the head-mounted displays

that were used. They recommend using technologies that allow participants to maintain their view of each other as well as the design representation (such as SAR or hand-held AR devices) as this enables communication through gestures and other non-verbal communication types.

Akaoka et al. (2010) investigated the use of SAR technology in student group design projects by introducing their 'DisplayObjects' system. The DisplayObjects workbench uses SAR technology to project graphics and user interface elements on to physical, low-cost prototypes. Testers liked the interactive, hands-on approach and the ability to change elements quickly. However, they did find problems with hand occlusions (which is when the digital overlay appears on top of the user's hand when handling the design representation) and found the preparation of the SAR virtual model to be very time consuming.

Several studies have investigated the potential use of AR/SAR technologies in design evaluation and usability studies. Porter et al. (2010) report on the testing of a SAR-based system that incorporates finger tracking. They compared the time taken to interact with buttons on the prototypes for groups using SAR and standard prototypes. Mean 'button-press time' was significantly increased in the SAR condition and one third reported that not having a physical button affected their interactions. Despite these limitations, many participants felt that SAR provided a good visual representation of the concept and 88% of participants agreed that SAR technology would be useful as a design tool.

Park and Moon (2013) identified a number of problems with the current range of AR technologies when they are applied in design evaluation studies. These include: hand occlusions; difficulty in supporting user interaction with the prototype (such as those identified in the Porter et al. (2010) study) at low cost and without creating an uncomfortable device; and the fact that many AR systems require a heavy, bulky headset to be worn by the user with a limited field of view. They present technical solutions to each of these problems with promising results from laboratory-based studies with students.

Irlitti and von Itzstein (2013) report on the development of the 'SARventor', which combines SAR with three tangible user interface 'tools'. The system was presented to three experts from architecture and industrial design. Challenges noted by the reviewers included the lack of a visible toolkit and the inability to manipulate volume (3D geometry). Despite this, the reviewers felt there was '…a strong case towards being used as a collaborative tool for use in feedback sessions between designers and stakeholders'.

These studies show there is significant potential for the application of AR and SAR technologies in applications such as group design (designers working together), evaluation studies and usability studies. However, it is noticeable that there is limited previous work on the use of AR and SAR technologies within the specific application of co-creative design sessions (designers working with non-designers). This is the application that is addressed by this study, as part of the SPARK project - which is introduced in the following section.

1.2. Background on the SPARK project and SAR technology

The SPARK project is a three-year, collaborative project that aims to develop and test a SAR platform for use in co-creative design sessions. The prototype technology has been developed and features: multi-projector visualisation; an infrared optical tracking system; a tablet PC-based user interface – shown in Figure 1 - that can be used to modify the digital overlay content; and an information system that is used for managing the digital assets and reviewing completed sessions. To prepare for a session, the designer must first create a basic physical model of the product to be worked on during the session. The model can be made using rapid prototyping or even constructed using cardboard, as long as it is white or light in colour. The 3D geometry data of the model is shared with the SAR system to enable accurate visualisation. Infra-red markers (or a marker constellation) are added to the physical model to enable tracking of the position and orientation of the model. Finally, the designer needs to upload digital assets that may be required during the session (e.g. logos, photographs, images etc.) to the information system. During the session, the designer can start with a blank model and add the digital overlay content element by element, or they can start with a pre-prepared concept that, which can then be modified during the session using the tablet PC-based user interface.
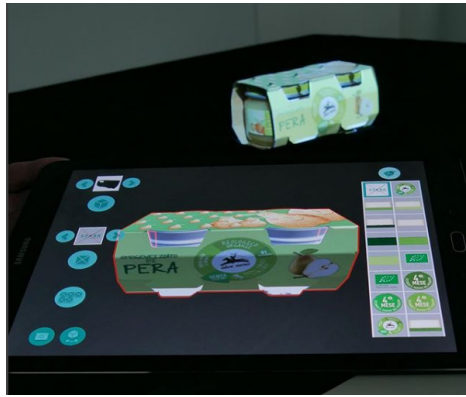
Figure 1. The SPARK tablet PC-based user interface (foreground) and SAR model (background)

2. Methodology

2.1. Description of experimental conditions

For the co-creative sessions three conditions were tested. The 'SAR condition' involved the use of the SPARK Spatial Augmented Reality technology, as described in Section 1.2. One designer used the SPARK user interface, installed on a tablet PC (Figure 1), to manipulate the SAR prototype based on the comments and suggestions from the other participants. The second designer and the end users viewed the SAR prototype to see the results of the manipulations.

The second condition was an 'AR condition', which involved the use of a physical prototype (prepared with optical tracking markers) and a tablet PC with an AR app installed. When viewed through the tablet PC, a digital overlay appeared on the physical prototype to create an augmented view – see Figure 2. One designer and the two end users shared a tablet PC to view the AR prototype, whilst the second designer used a separate tablet PC to manipulate the SPARK user interface (as in the SAR condition).
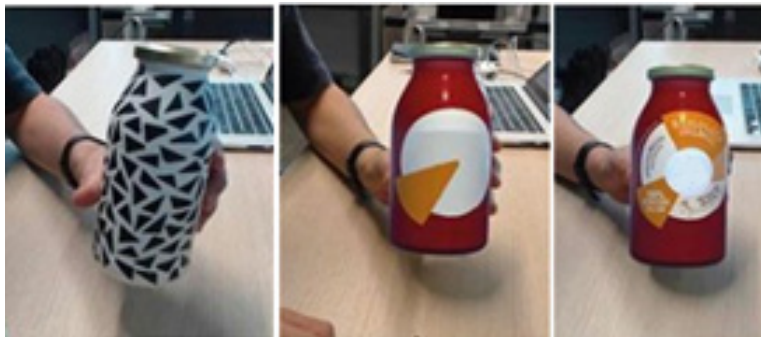

Figure 2. View of the augmented reality prototype i) without augmentation, showing the marker pattern ii) with augmentation - design in progress, iii) with augmentation - completed design proposal

For the third condition, the designers were asked to use conventional materials and tools to prepare the design representations for their sessions (henceforth, we refer to this as the 'conventional' condition). For the Stimulo session, the initial designer proposals were displayed on a large television screen using presentation software. After this, physical prototypes featuring neutral colours were presented and Pantone colour swatches were used to discuss alternative colour schemes - see Figure 3 (left). For the Artefice conventional session, the designers elected to use a collage method, which involved pre-preparing a variety of logos and graphic elements as stickers that could be applied to the cardboard sleeve of the soup packaging, re-positioned as required and further elements added by hand drawing directly on to the cardboard sleeve - see Figure 3 (right).

Figure 3. Left - Physical models and Pantone© colour matching system used in the Stimulo 'conventional' session; Right - collage system used in the Artefice 'conventional' session

2.2. Case studies and participants

The experiments featured designers from Stimulo and Artefice, with three sessions organised with each industrial partner i.e. six sessions in total – summarised in Table 1. For the Artefice sessions it was possible to have different teams of designers to work on the same product and same initial brief for each of the three conditions. For the sessions with Stimulo, only two designers were available to participate in the experiments. With this limitation, using the same product and brief for each condition was not desirable as it would have risked the designers (consciously or sub-consciously) carrying over ideas from one session to another and becoming bored of the task due to the repetition. It was therefore necessary to vary the case study product for each of the conditions, although efforts were made to ensure that the case study products and session briefs were as similar as possible in scope, task and design stage.

Table 1. Summary of the products, scope and participants for each condition

		SAR condition	AR condition	Conventional condition
Stimulo	Product description	Hand held device for assessment of human exposure to electromagnetic fields	Smart fitness product to monitor performance when using gym equipment	Hand held device for communicating your location in an emergency
	Session brief	Define the colours, materials and finish of the main housing. Define the location and pattern of LED status lights and speaker. Location of logo.	Define the colours, materials and finish of the main housing. Location of logo.	Define the colours, materials and finish of the main housing for specific environments. Define the location and pattern of LED status lights.
	End users	Female, age 18-30 Male, age 18-30 Male, age 45-60	Female, age 18-30 Male, age 18-30 Male, age 45-60	Female, age 18-30 Male, age 18-30 Female, age 45-60
	Designers	Creative Director, 14 years of experience, male Designer and Business Developer, 15 years of experience, male		
Artefice	Product description	Fresh soup - single serving in plastic bowl with film lid and cardboard sleeve		
	Session brief	Further develop three pre-prepared alternative designs for the cardboard sleeve graphics and layout by combining graphical elements (colours, logos, text, images etc.) in order to propose a complete packaging design.		
	End users	Female, age 18-30 Male, age 18-30	Male, age 18-30 Male, age 18-30	Female, age 30-45 Female, age 30-45
	Designers	Digital Creative Director, 16 years of experience, female Art Director, 18 years of experience, female	Senior Art Director, 19 years of experience, male Graphic Designer, 10 years of experience, male	Art Director, 10 years of experience, female Junior Art Director, 1 year of experience, female

The teams of designers for each session were selected to ensure reasonable consistency across the conditions in terms of experience and skills. The end users were selected to match the target demographic of the case study product for the session they would participate in i.e. '18-60 year old adults who enjoy hiking in the mountains' for the Stimulo emergency location beacon session.

2.3. Data collection and analysis protocol

Three main research methods were employed, linked to the three research questions stated in Section 1. To test if AR/SAR-based design representations result in more effective sessions (RQ1), a suite of quantitative co-creative session performance metrics were applied. To establish if designers perceive AR-SAR-based design representations to be more effective than conventional design representations (RQ2), the Creativity Support Index (Cherry and Latulipe, 2014) method was employed. Finally, to understand the specific strengths and weaknesses of AR-SAR-based design representations in comparison with conventional design representations (RQ3), a follow-up survey was conducted with the designers that participated in the experiments. The data collection and analysis protocol for each of these methods are presented in the following sub-sections.

2.3.1. Application of the co-creative session performance metrics

A suite of metrics suitable for the analysis of co-creative sessions has previously been developed (O'Hare et al., 2016) and later refined (Dekoninck et al., 2018). Those co-creative session performance metrics are presented in Table 2.

Table 2. Definition the co-creative session performance metrics

Metric title	Metric definition
Quantity of ideas	Quantity of ideas generated during the session, counted as the number of screenshots taken by participants, and subsequently verified in the post-session interview.
Variety of ideas	Variety (original) - Number of original feature rows that contain a new idea counted on the morphological chart created by an observer in the session.
	Variety (new) - Number of new feature rows added to the morphological chart created by an 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 designers in the post-session interview.
Novelty of ideas	Mean average score for novelty from each participant in the post-session interview for each of the ideas captured as a screenshot during the session.
Task Progress	Task Progress = 3pts x (Number of high importance tasks resolved or 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 rejected ideas*/(Quantity of ideas [as per the definition above] - Desired number of ideas to retain)

*The 'number of rejected ideas' represents the number of ideas that were generated during the session but were eliminated from the design process by the designers during the post-session interview

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; ideas that 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 (which is the 'Desired number of ideas to retain' used in the Filtering Effectiveness metric).

DESIGN SUPPORT TOOLS

At the start of the co-creative session itself, all participants were asked to request a screenshot/picture to be taken whenever they felt that they 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 in to the SPARK platform user interface, or by taking a photograph with their cameraphone during the conventional sessions.

A researcher 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 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.

After the session, a joint interview with both of the designers from 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 were duplicates or captured by mistake (for the Quantity metric). The designers were 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. Such rows were counted as 'new rows' for the Variety metric, whilst all other rows not identified as 'new rows' were counted as 'original rows'.

Finally, the list of open tasks from pre-session interview was revisited to check which tasks had been completed and what new tasks, if any, had been generated during the session.

2.3.2. Application of the Creativity Support Index survey

The Creativity Support Index (CSI) survey was developed by Cherry and Latulipe (2014) to help evaluate the support for creativity provided by ICT tools. The CSI survey was selected because it is based on the well-known NASA Task Load Index survey (Hart and Staveland, 1988) but with more focus on evaluating the support provided by ICT tools for creative activities, and so it is well aligned with the objectives of the SPARK platform.

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.

The CSI survey was generally administered immediately after the session, as part of the post-session interview. Only the designer that had manipulated the user interface was required to complete the survey from the AR and SAR conditions, whilst all designers completed the survey separately for the conventional sessions.

2.3.3. Follow-up survey

A follow-up survey was sent to the designers via email after the session. It featured four questions that were intended to capture qualitative feedback from the designers about the performance of the session. The four questions were:

- What were all the things that went well during the session?
- How did the [SAR/AR/other] tool you were using contribute to the positive aspects of the session you have described above?
- What were all the things that were challenging about the session?
- How did the [SAR/AR/other] tool you were using contribute to the challenging aspects of the session you have described above?

3. Results

3.1. Results of co-creation session performance metrics

Table 3 presents the results of the application of the co-creative performance metrics (Dekoninck et al., 2018) to the six co-creative sessions completed. The best score (or joint best score) from the three conditions is highlighted in bold for both the Stimulo and Artefice sessions.

Table 3. Results of the co-creative performance metrics application

Metric title		Stimulo			Artefice		
		SAR	AR	Conventional	SAR	AR	Conventional
Quantity of ideas		**8**	**8**	6	**11**	4	5
Variety of ideas	Original	**5**	1	4	2	4	**5**
	New	**1**	**1**	**1**	0	0	**1**
Quality of ideas		=8 – 4 rejected = 4	= 8 - 3 rejected = **5**	= 6 – 5 rejected = 1	= 11 – 7 rejected = **3**	= 4 – 3 rejected = 1	= 5 – 3 rejected = 2
Novelty of ideas		= 44 ÷ 8 = 5.5	= 51 ÷ 8 = **6.4**	= 23 ÷ 6 = 3.83	= 7 ÷ 3 = 2.3	= 9 ÷ 4 = 2.3	= 19 ÷ 5 = **3.8**
Task progress		1xHigh Total = 3	2xHigh 1xMed **Total = 8**	1xMed Total = 2	2xHigh **Total = 6**	1xHigh Total = 3	1xHigh 1xMed 1xLow **Total = 6**
Filtering Effectiveness		= 4/(8-1) = 0.57	= 3/(8-5) = 1	= 5/(6-1) = 1	= 8/(11-1) = 0.8	= 3/(4-2) = 1.5	= 3/(5-3) = 1.5

For the Stimulo sessions, the SAR and AR condition performed best or joint best against the idea generation metrics (quantity, variety, quality, and novelty of ideas), task progress and filtering effectiveness metrics, with the AR condition generally offering the best performance. The novelty and quality of ideas stand out as the particular successes for the SAR and AR conditions over the conventional condition at Stimulo. For the Artefice sessions, the conventional condition performed best or joint best in terms of the variety, novelty, task progress and filtering effectiveness metrics, but was not so good on the quantity and quality metrics. Overall, the results of the metrics application present a mixed picture, with none of the conditions consistently outperforming the others across all metrics and both companies.

3.2. Results of Creativity Support Index assessment

Table 4 presents the results of the CSI assessment. The average score is presented by condition and with the breakdown by factor. The maximum score for each cell is 10. The overall CSI score is provided in the final row and is the average from across the Stimulo and Artefice designers that participated in the same condition. The maximum overall CSI score is 100 and scores above 90 indicate that the tool offers excellent support and scores below 50 indicate poor support (Cherry and Latulipe, 2014). On this basis, both the SAR and AR performed reasonably well, whereas the conventional condition was considered poor. The 'average count' column is an indicator of the relative importance of each of the factors. The maximum score is five. The 'collaboration' aspect had the highest average count (4.1), followed by 'exploration' (3.7) and 'immersion' (3.0). The AR condition performs consistently well against these three most important aspects. The SAR condition scored well on the collaboration and exploration

aspects but very poorly on the immersion aspect – probably caused by some technical difficulties experienced during those sessions.

Table 4. Results of the Creativity Support Index assessment

Aspect	Average Score			Average count (all conditions)
	SAR	AR	Conventional	
Collaboration	7.5	8.3	5.8	4.1
Exploration	7.0	7.3	4.0	3.7
Immersion	2.8	7.5	4.3	3.0
Expressiveness	4.5	5.0	4.5	2.0
Results worth the effort	7.8	6.3	5.3	1.3
Enjoyment	9.0	7.5	5.0	0.9
Overall CSI score	67.2	74.7	49.0	

3.3. Results of follow-up survey

Table 5 provides a summary of the key points from the designer follow-up survey. The points were summarised by the researcher from the original responses provided by the designers.

Table 5. Summary of designer feedback from follow-up survey

Question	Condition		
	SAR	AR	Conventional
What went well?	*Artefice:* Helps with detailed refinements. Helps to quickly rule out poor suggestions from end users. *Stimulo:* Freedom to try many different ideas.	*Artefice:* Improved interaction with end users. *Stimulo:* Good interaction/ communication with end users.	*Artefice:* Good empathy with the end users, who were willing and able to provide good input. *Stimulo:* End users were positive and focused.
How did tool contribute to positive aspects?	*Artefice:* Quick, real-time modification of a tangible prototype facilitates co-creation. *Stimulo:* Able to generate and test some new ideas for colours and logo position.	*Artefice:* Real-time modification improved interaction with the end users. *Stimulo:* Intuitive sharing of ideas between end users and designer allowed quick iteration of concept.	*Artefice:* Intuitive interaction method that enabled the end users to participate in an uninhibited manner. *Stimulo:* It was a basic way to support the engagement between the end users and the designer.
What was challenging?	*Artefice:* Some end users frustrated by perceived limitations of the system. *Stimulo:* Technical problems had an impact. Limited interaction with the prototype. Major differences between the designer's view (tablet) and end users' view (SAR).	*Artefice:* Various technical limitations and failures hindered and disrupted the session. *Stimulo:* More chaotic, less focused session with more random/trial and error - requires more pro-active facilitation.	*Artefice:* Limited range of elements and hand drawn elements limits the quality/realism/ fidelity of the final outcome. *Stimulo:* LED position options difficult to represent.
How did tool contribute to challenging aspects?	*Artefice:* Inability to add new assets during the session or modify standard elements limited the freedom of the session. SAR model lacks realism. *Stimulo:* 3D effect and on-the-fly changes within SAR is useful but need more user interaction features.	*Artefice:* Technical problems caused disruption. Limitations of the system discouraged the end users. *Stimulo:* Tablet create a barrier to direct interaction with the real prototype.	*Artefice:* Limited range of pre-prepared elements meant that the final outcome did not entirely represent what was desired/discussed. *Stimulo:* Designer had to build a fourth concept (mixing elements of the three pre-prepared concepts).

4. Discussion

The first research question asked: "Do co-creative sessions involving AR/SAR-based design representations result in more effective sessions in terms of idea generation, progress on design tasks, and filtering of ideas than those conducted with conventional design representations?" The results of the co-creative performance metrics have highlighted certain aspects, such as the quality and novelty of ideas within the Stimulo sessions, where the AR/SAR conditions significantly outperformed the conventional design representations condition. Technical faults during the SAR session with Stimulo and during the AR session with Artefice appear to have affected the performance of these sessions, as confirmed by the designers who explicitly mentioned these technical faults in the follow-up survey. Given the impact of these technical faults, further experimental sessions are required in order to be able to provide a definitive response to this research question. Furthermore, it might not be possible to provide a single response to this question, as the effectiveness of AR/SAR-based design representations is likely to vary significantly depending on the nature of the design task to be completed and the type of participants. This will be an important consideration for future experiments.

The second research question asked: "Do designers perceive AR/SAR-based design representations to be more effective for co-creative design sessions than conventional design representations?" The results of the CSI assessment suggest that designers do indeed perceive AR/SAR-based design representations to be more effective for co-creative design sessions than conventional design representations. However, there are clearly areas for improvement, such as the sense of immersion in the tool when using the SAR technology. Hence the tentative conclusion is that designers do indeed perceive AR-SAR-based design representations to be more effective, but this will need further experimental sessions to be confirmed.

The third research question asked: "What are the specific strengths and weaknesses of AR/SAR-based design representations in comparison with conventional design representations?" The main perceived strengths of both the AR and SAR conditions was the enhanced collaboration with the end users and the ability to make quick, real-time modifications to the prototype to try out new ideas. A notable example of this came in the SAR condition session with Artefice, in which the end user made a proposal concerning the position of a logo. The designer was confident that this proposal would not enhance the design. In the follow-up survey she noted that with the SAR technology it was quick and simple to modify the prototype according to the proposal, show it to the end user and get them to agree that it was not a good proposal before reverting to the original logo position. This type of idea elimination activity can be particularly helpful in co-creative design sessions in which clients are the collaborator.

It terms of weaknesses, the reliability/robustness of the AR and SAR systems was a problem for both companies, with designers commenting that the technical faults that occurred during the sessions resulted in awkward pauses that disrupted the natural flow of the session and caused frustration for all participants. Some specific technical limitations were also identified. From Artefice's perspective, the SAR prototype lacked realism because it was too bright/reflective. For them, the AR condition was preferred as the quality and fidelity of the visualisation was considered to be better. They also noted that the inability to add new graphic elements from external sources during the session - a limitation of the current SPARK Platform - caused frustration as it meant that they were not able to respond to some suggestions by the end users. From Stimulo's perspective, a key limitation was the significant differences in colour hue and shade that were noted when viewing the design representation on the designer's user interface compared to the view seen by the other participants when looking either at the SAR prototype or the AR prototype through the tablet. Given that selecting colours, materials and finishes was a key objective of the Stimulo sessions, this was a major problem. Another challenge identified by Stimulo was that the flexibility of the AR and SAR systems reduced the control that they had in managing the session, as there were "…no boundaries for the end users to participate and share ideas." Whilst this may be seen as a positive aspect in terms of end user involvement and engagement, Stimulo felt that this enthusiasm from the end users to contribute their ideas needed more focus to be productive. Their reflection was that, in future sessions using AR/SAR-based design representations, the designers would need to be more pro-active in their facilitation of the session to keep it on track and avoid too much unproductive trial-and-error experimentation.

5. Conclusions

The objectives of this study were: to begin gathering data comparing the effectiveness of co-creative design sessions conducted with AR/SAR-based design representations versus conventional design representations; and to identify specific strengths and weaknesses of AR/SAR-based design representations in comparison with conventional design representations. These objectives led to the formulation of three research questions, which were addressed through controlled experiments with professional designers working on real projects with end users.

It was not possible from this small-scale study to provide a definite response to the first research question concerning the effectiveness of using AR/SAR-based design representations within co-creative design sessions. Nonetheless, some aspects, such as the quality and novelty of ideas within the Stimulo sessions, were found to be significantly better using the AR/SAR-based design representations compared to the conventional design representations condition. Of course, care must be taken to avoid bias in the interpretation of the results, as both the researchers and the designers from the industry partners have invested significant time and effort in the project based on the assumption that AR/SAR-based design representations will result in more effective co-creative design sessions. To mitigate this risk of bias, we do not propose a firm conclusion on this question until later in the project, when we will have a larger data set completed, including sessions completed with designers from outside of the SPARK consortium.

The second research question focused on the designers' perception of the effectiveness of AR/SAR-based design representations for co-creative design sessions. The evidence from the CSI assessment suggested that designers do perceive AR/SAR-based design representations to be more effective, but this will need further experimental sessions to be confirmed.

The third research question was intended to help understand the relative strengths and weaknesses of AR and SAR-based design representations in comparison to conventional design representations. The main perceived strengths of both the AR and SAR conditions was the enhanced collaboration with the end users and the ability to make quick, real-time modifications to the prototype to try out new ideas. Weaknesses identified included reliability issues, the overall insufficient quality of the SAR visualisation, and problems in both the AR and SAR conditions with differences in colour rendering.

Within the SPARK project, the next steps will be to address the limitations listed above before conducting further trials with designers and student groups. In addition, a specific request from several designers was to enhance the system to better support user interaction design. For example, the ability to represent how lights on the product will flash or pulse on and off in response to user actions. Similarly, with sound, designers would like the ability to test different types of notification or warning sounds with end users during the session. From these suggestions, we can conclude that the full potential of SAR/AR-based design representations has not yet been exploited, particularly in terms of interaction design. Nonetheless, these experiments have provided quantitative and qualitative evidence to suggest that AR/SAR-based design representations can help to improve certain aspects of co-creative design sessions and have helped us to identify some of the relative strengths and weaknesses of the technology within this application. These insights will be of interest to AR/SAR technology developers as well as to researchers investigating co-creative design activities.

Acknowledgment

References

Akaoka, E., Ginn, T. and Vertegaal, R. (2010), "DisplayObjects: Prototyping Functional Physical Interfaces on 3D Styrofoam, Paper or Cardboard Models", *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction - TEI '10*, pp. 49-56. https://doi.org/10.1145/1709886.1709897

Billinghurst, M., Belcher, D., Gupta, A. and Kiyokawa, K. (2003), "Communication Behaviors in Colocated Collaborative AR Interfaces", *International Journal of Human-Computer Interaction*, Vol. 16 No. 3, pp. 395–423. https://doi.org/10.1207/S15327590IJHC1603_2

Bodker, S. (1998), "Understanding Representation in Design", *Human–Computer Interaction*, Vol. 13 No. 2, pp. 107–125. https://doi.org/10.1207/s15327051hci1302_1

Business Innovation Observatory (2014), Design for Innovation - Co-creation design as a new way of value creation, European Union, Brussels.

Cherry, E. and Latulipe, C. (2014), "Quantifying the Creativity Support of Digital Tools through the Creativity Support Index", *ACM Transactions on Computer-Human Interaction*, Vol. 21 No. 4, pp. 1–25. https://doi.org/10.1145/2617588

Dekoninck, E.A., O'Hare, J., Giunta, L., Masclet, C. and Cascini, G. (2018), "Exploring ways to speed up the application of metrics to assess co-creative design sessions", *Proceedings of the Fifth International Conference on Design Creativity*, The Design Society, Bath.

Hart, S.G. and Staveland, L.E. (1988), "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research", *Advances in Psychology*, Vol. 52, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Irlitti, A. and von Itzstein, S. (2013), "Validating constraint driven design techniques in spatial augmented reality", *Proceedings of the Fourteenth Australasian User Interface Conference, AUIC, Vol. 139*, pp. 63–72.

O'Hare, J. Dekoninck E.A., Mombeshora, M. and Cascini, G. (2016), "Developing metrics to assess technology-enabled creative co-design sessions", *Proceedings of the 4th International Conference on Design Creativity (ICDC)*, The Design Society, Glasgow.

Park, H. and Moon, H.C. (2013), "Design evaluation of information appliances using augmented reality-based tangible interaction", *Computers in Industry*, Vol. 64 No. 7, pp. 854–868. https://doi.org/10.1016/j.compind.2013.05.006

Porter, S.R., Marner, M., Smith, R., Zucco, J. and Thomas, B. (2010), "Validating spatial augmented reality for interactive rapid prototyping", *Proceedings of the 9th IEEE International Symposium on Mixed and Augmented Reality 2010: Science and Technology, IEEE*, pp. 265–266. https://doi.org/10.1109/ISMAR.2010.5643599

Sanders, E.B.-N. and Stappers, P.J. (2014), "Probes, toolkits and prototypes: three approaches to making in codesigning", *CoDesign*, Vol. 10 No. 1, pp. 5–14. https://doi.org/10.1080/15710882.2014.888183

Stacey, M. and Eckert, C. (2003), "Against ambiguity", *Computer Supported Cooperative Work*, Vol. 12 No. 2, pp. 153–183. https://doi.org/10.1023/A:1023924110279

Dr. Jamie Alexander O'Hare, Research Associate
University of Bath, Department of Mechanical Engineering
Department of Mechanical Engineering, BA2 7AY Bath, United Kingdom
Email: en0jao@bath.ac.uk