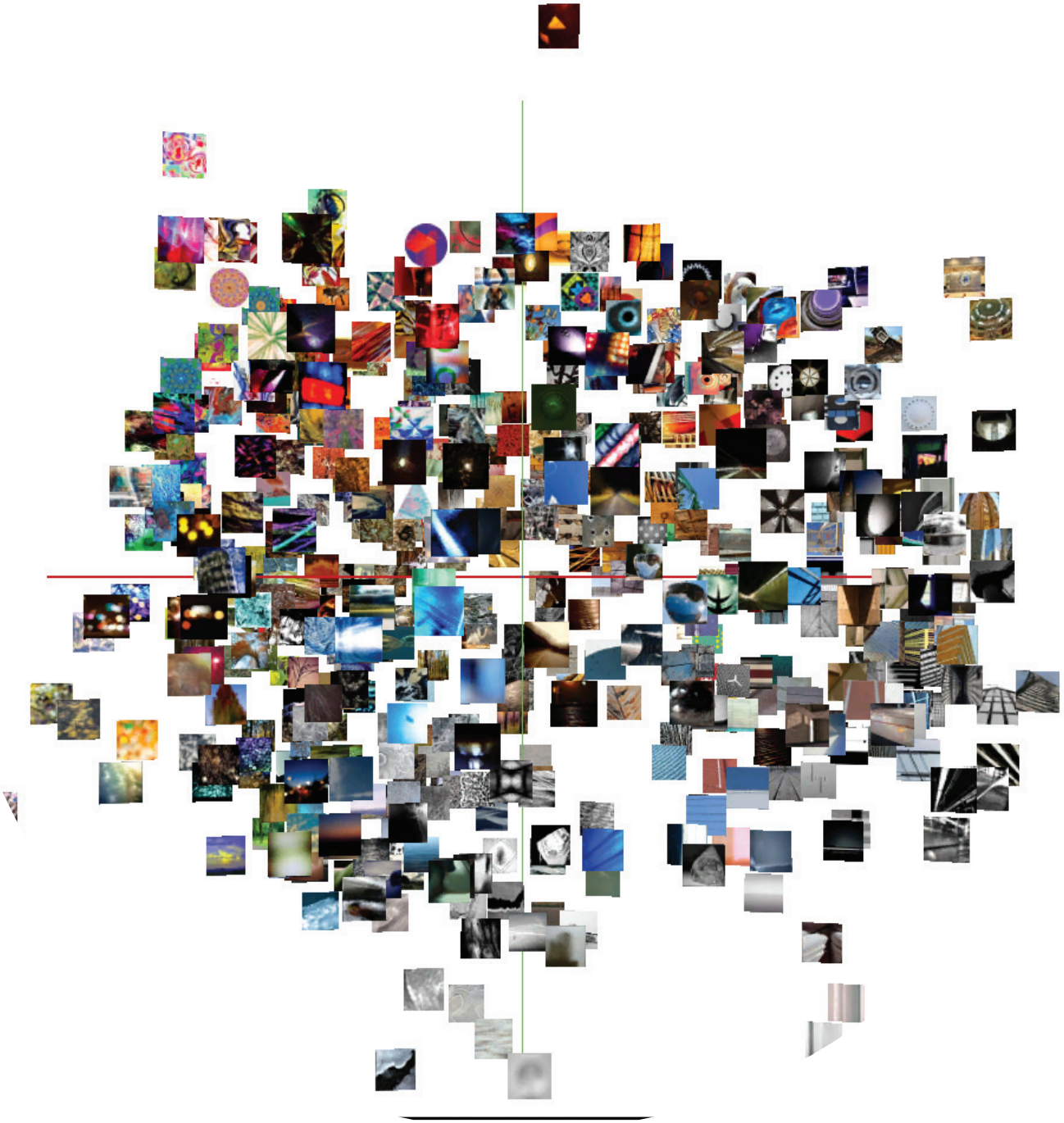
DR BRITTA KALKREUTER, DAVID ROBB,
PROF MIKE CHANTLER, DR STEFANO PADILLA
HERIOT-WATT UNIVERSITY



MANAGING CREATIVE CONVERSATIONS BETWEEN DESIGNERS AND CONSUMERS

CREATIVE COMMUNICATIONS | CO-DESIGN | FASHION | SEMIOTICS OF DESIGN

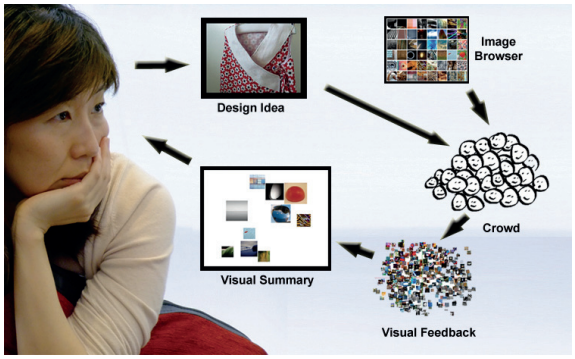
ABSTRACT

THE PAPER DESCRIBES ASPECTS OF A COLLABORATIVE PROJECT WHICH BRINGS TOGETHER RESEARCHERS FROM THE SCHOOL OF MATHEMATICAL AND COMPUTER SCIENCES AND THE SCHOOL OF TEXTILES AND DESIGN AT HERIOT-WATT UNIVERSITY.

The project is funded as part of Heriot-Watt's Creativity, Design and Innovation (CDI) initiative and combines computational expertise with design thinking to investigate how the growing appetite for consumer participation in fashion design can be enabled and enriched by facilitating web based creative conversations between crowds and designers. The project concentrates on visual forms of communication so as to mirror the highly visual nature of fashion design inspiration. The paper gives an overview of the project design to date before presenting key findings on the communicative certainty of visuals based on two experiments:

The first experiment tested the level of agreement amongst distinct demographic groups when asked to assign visuals from a closed data set of 500 abstract images to 20 terms chosen for being descriptive of material (e.g. textured) and emotive qualities (e.g. enjoyment, pleasure) of fashion. The second test explored the communicative certainty of visuals by asking a fresh set of observers to relate the selected visuals back to the original terms. The resulting data advances our technical knowledge of how mass visual feedback can best be aggregated to convey commentary from a consumer to a designer.

FOR REASONS RANGING FROM SUSTAINABILITY TO A BACKLASH AGAINST THE STANDARDISATION OF FAST FASHION, AN INCREASING NUMBER OF WEBSITES NOW INVITE CONSUMERS TO BECOME INVOLVED IN THE DESIGN OF FASHION, MOSTLY IN THE FORM OF CUSTOMISATION.



1

The comparative aspects of the second experiment which analysed the performance of raw selection data against their visual aggregations have indicated that a scaling up of the project is possible as the communicative certainty of showing all chosen images was found to be broadly equal to showing condensed collages of image choices. With this in mind, the project is now moving towards the challenge of how best to utilise the developed interfaces and algorithms to enable crowds and designers to remotely take part in co-design cycles. The project has also highlighted opportunities for the semiotic study of fashion intelligence and the visual communication potential of design itself.

INTRODUCTION AND CONTEXT

The aim of the project is to connect consumers and designer in a creative process and was nicknamed HeadCrowd early on because it seeks to connect the consuming crowd with the mind of the designer (see figure 1). While fashion design is a discipline where many consumers become creators on a daily basis when selecting and styling their outfits, actual co-design in fashion is rather unusual as the designer's mind might use market intelligence to second guess (or fuel) the desires of the consuming crowd but will rarely engage with it on the actual creation of designs. This is unlike disciplines like architecture where routine customisation and stake holder involvement has led to a much more standardised participatory approach between client and designer.

For reasons ranging from sustainability to a backlash against the standardisation of fast fashion, an increasing number of websites now invite consumers to become involved in the design of fashion, mostly in the form of customisation, and requiring the consumer to select from a set of pre-determined variations available for example companies such as Blank Label and United Styles. Our project seeks to go beyond customisation by investigating how rich and mobile web applications can be employed to best facilitate visual feedback between designers and co-creating crowds. The idea originated in discussions on Creativity, Design and Innovation between the Schools of Mathematical and Computer Sciences and the School of Textiles and Design at Heriot-Watt University, where colleagues explored how recent advances in the computing of visual, and specifically textile related data could be harnessed for the increasingly popular field of participatory design in fashion The project acknowledges a change in communication conventions brought about by the meteoric rise of handheld mobile devices: it has been found that users of touch screens tend to type less than PC users when called on to enter input, and HeadCrowd's development of touch operated visual feedback selection seeks to allow mobile users to have a differentiated communication through means suitable for their chosen interface.

The project is also driven by the realisation that much of the so called co-creational activity in fashion relies almost solely on textual feedback from consumers, and that much of this activity is currently concentrated

towards the final phases of the design process, making it a tool solely for sales tactics, branding and mass customisation.

In contrast, HeadCrowd seeks to acknowledge David Crow's 'cultural shift from words to pictures' (Crow 2006), to facilitate early and iterative creative communication between professionals and users to achieve co-creation throughout a process that leads to a product unknown at the start of engagement. In the spirit of the Cluetrain manifesto, we start by believing that 'markets are conversations'. (Levine, Locke, Searls & Weinberger 1999). Initially understood as a wake-up call for global companies to heed the voice of ever more vocal and connected consumers, the Cluetrain phenomenon asks the professional designer to acknowledge that their customers are increasingly seeking not to just satisfy their consumerist, but their creative needs as well.

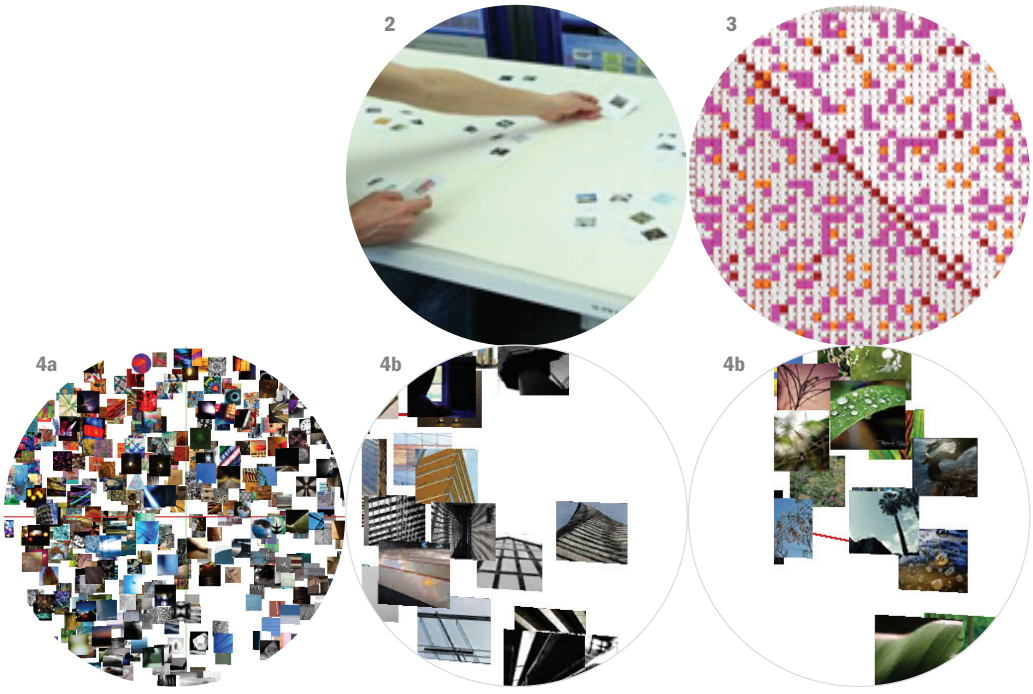
Aware of the potential implications of this development for creative professionals, the project is working closely with staff and students in the School of Textiles and Design at Heriot-Watt University to gauge reactions and evaluate concerns throughout.

A final research aim of HeadCrowd is to use mass computational means to generate sufficient data for semiotic analysis of the visual feedback quality, and in order to create a platform that is able to reach crowds and compute feedback from them. Specific attention will be paid to creating and testing methods of generating perceptually relevant visual précis of mass image selections in order to facilitate manageable and effective communication between large numbers of participants.

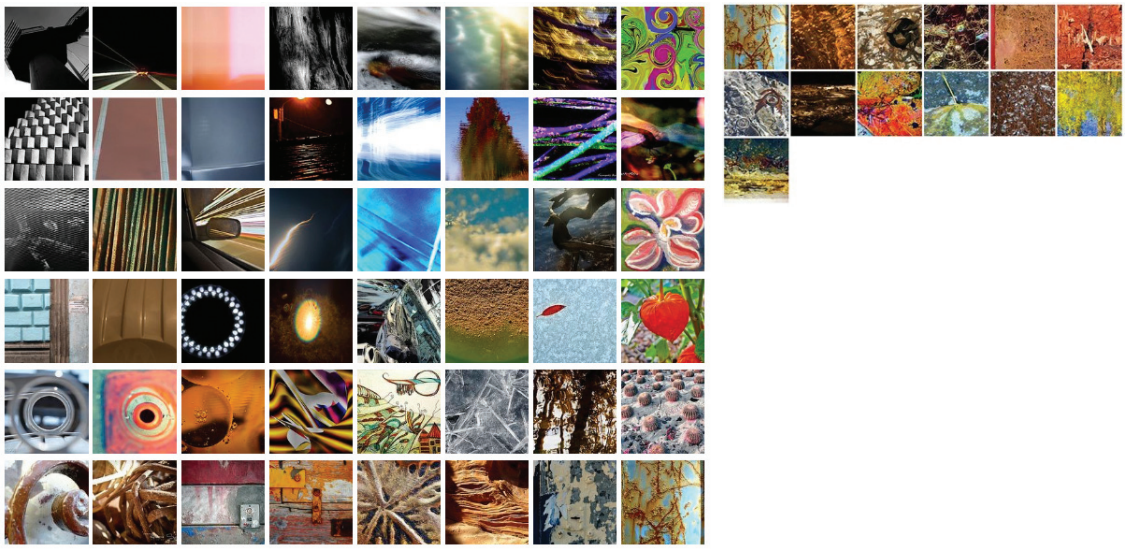
FEEDBACK PHASE 1: CHOOSING IMAGES TO DESCRIBE TERMS

The first technical phase of the project sought to develop prototype interfaces that would allow fast and intuitive navigation of visuals by remote crowds in order to gain research data on image selection behaviour: this essentially meant developing algorithms and protocols which automate the organisation of large rich data sets and facilitate selection from this set without predetermining choices.

In order to enable and study visual communication behaviour across non-design and design demographics, keeping in mind the eventual target audience for the co-creation process, a closed image set was chosen, so that image verification rather than image creation was demanded of participants. In a similar bid to



5



minimise prejudice on the basis of the varied cultural background of participants, the images for the initial set were chosen for their non-representational and non-symbolic content, initially by targeting images tagged as abstract on Flickr, then by computational and manual elimination of erroneously tagged visuals, and to reject duplicates. Some 2000 images were initially chosen from Flickr, because its many 'Creative Commons' licenses avoid copyright issues, and were subsequently reduced to 500 which would form the visual vocabulary for the experiment. Making this intentionally large vocabulary manageable for use required a similarity matrix that was recognised by a computer, though its foundation was deliberately based on the perception of humans who would seek to communicate with them.

The study used the tightly controlled settings suggested by Halley, (2011) to sort an initial number of 100 images by 20 volunteers from a fashion demographic in a face to face exercise (figure 2) so that signs of fatigue and loss of interest could be controlled.

These 100 images were augmented by the remaining 400 images using (remote but human) participants from Amazon Mechanical Turk (MTurk), again following Halley's tested method. The collected similarity judgments were assembled into a similarity matrix relating each image to every other image, see figure 3.

The relative similarity or dissimilarity of images within the set was visualised by using dimensionality reduction and viewing the images in a 3D space where similar images are close to each other and dissimilar images are further apart. Themed clusters of images were noticeable, see figure 4.

For presentation to users in the crowd, the similarity information was used to organise the 500 images into 48 similarity stacks as these could be comfortably displayed on an iPad, thus meeting the requirement for convenient and mobile use of the visual selection interface. The organisation of the stacks follows a self-organising map (SOM) layout, see figure 5.

Due to the way in which card sorting and MTurk delivered complex and multi-relational similarity judgements on the 500 images, the vocabulary could be arranged in any number of stacks. This would make the immediately visible number of top level images greater or smaller, and reduce or increase the number of images beneath this top level interface surface accordingly. Being able to change the display parameters for the 500 images in this way offers the potential of using a wider range of mobile devices, yet is likely to have implications for the selection behaviour of observers. Any changes to the image display interface will therefore have to be carefully considered in the data evaluation.

With a large visual vocabulary now available on an easily navigated interface, HeadCrowd proceeded to test the communicative value of the vocabulary by asking observers to select 3 images each, for a total of 20 textual terms chosen for their potential significance in the design process of fashion and textiles: 10 terms were based on Methven's research into textural qualities of cloth (Methven et al 2011), and 10 further terms were chosen from the Geneva Emotional

BRITTLE	ASTONISHMENT, SURPRISE
COARSE	DISGUST, REPULSION
CRUMPLING	EMBARRASSMENT, SHAME
DELICATE	ENJOYMENT, PLEASURE
FUZZY	INVOLVEMENT, INTEREST
FLEXIBLE	IRRITATION, ANGER
SMOOTH	SADNESS, DESPAIR
SOLID	TENDERNESS, FEELING LOVE
STICKY	WONDERMENT, FEELING AWE
TEXTURED	WORRY, FEAR

6

Wheel, (Scherer 2005), in recognition of fashion's non-material impact and qualities, see figure 6.

Effectively asking observers to translate text into images in this first experiment allowed the research to investigate the communicative certainty of the images, in other words to gauge to what extent Saussure's signifier (the image) is overlapping with the signified (textual term), (Chandler 2002). A comprehensive analysis of the limitations and results of this so called Fb1 experiment was published by Kalkreuter & Robb in the Nordic Journal and from a semiotic and sociological point of view, the results can be summarised as follows:

The 20 observers for the Fb1 pilot were drawn from a demographic that comprised 10 female and 10 male participants, 10 designers and 10 non-designers, with a mix of European and Asian ethnicities. A total of 1200 images (3 images for each of the 20 terms by each of the 20 observers) were recorded on a central server enabling quantitative analysis of the whole set of results as well as by personal demographic data relating to gender and professional background, and by type of term.

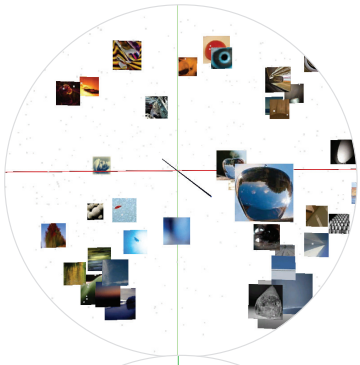
62.42% of all 1200 choices were singular selections, with just one fifth of images being chosen twice for the same term, and only 8.75% and 5.67% of selections for a term being a shared choice by 3 and 4 participants respectively, indicating an overall low strength of shared communicative value, describing a high level of ambivalence of the images' meaning instead.

Interestingly, however, the percentage of singular choices was significantly lower for material terms (55.33%) than it was for emotional terms (69.50%). Considering the fact that almost half of the choices when faced with picking just three out of 500 available images were agreed on by at least two of the 20 observers when it came to textural descriptions can be considered as significant in terms of agreed visual description of a texture. This trend is strongest for the term "solid" with a percentage of just 40% for singular choices and an impressive 60% for agreement between at least two observers on an image from this term, and weakest for the term 'involvement/interest', where every single image choice remained individual to just one observer (i.e. 60 separate images were suggested as a translation of that term).

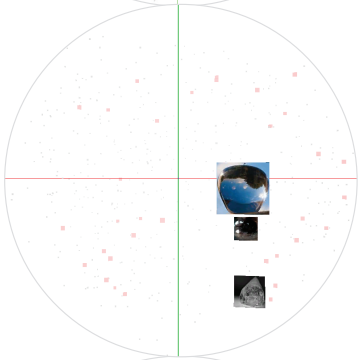
Regarding the pilot results according to the demographic data of gender and professional background makes interesting reading with regards to the study's context of co-creation in the fashion industry: Across all terms there is a higher than average disagreement on image selection within the male and within the female 'crowd' participating in the experiment, with 79.17% and 74.33% of singular choices respectively, and also a higher than average instance of individual image choices when considering separately the design and non-design crowds, with a shared instance of singular choices of 77.17% for each crowd, now only drawing data from 10 rather than 20 observers each, however. The results for singular choice, or less agreement, remain above the overall 20 observer average when considering emotional and textural terms separately. (Non-design textural 72.67% and design textural 73% as opposed to 55.33% of singular choices across all 20 observers)

The former result seems significant as fashion tends to be gender specific when looking at product but frequently involves different genders on the professional designer and consumer side. The latter seems to suggest no increased agreement on the communicative value of a given set of visuals amongst design professionals and could therefore be read as an endorsement of the Cluetrain manifesto and subsequently Lawson (Levine, Locke, Searls & Weinberger 1999) who acknowledge the creative potential of users and amateur designers.

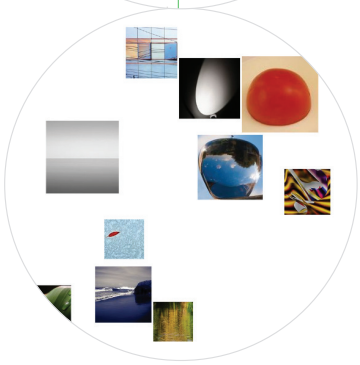
7



8



9



FEEDBACK PHASE 2: CONSTRUCTION OF SUMMARY COLLAGES

The second phase of the visual feedback experiments sought to create data in two distinct areas: firstly, to investigate the communicative value of the visual feedback given in phase 1, and secondly to analyse which feedback format would best facilitate a creative communication between observers. Feedback phase 2 therefore tested the selection behaviour of observers using the full 60 images selected by 20 observers for each term against the selection behaviour using summary collages of ten perceptually representative images derived from these 60 strong sets. The same similarity data used to produce the perceptually relevant image browser for phase 1 was used to construct the summary collages. The portion of the similarity data that applied to the images selected for each term in Fb1 was subjected to cluster analysis. This split each image set up into ten clusters based on the images' perceived similarity. Using multidimensional scaling (MDS) it is possible to visualise the images in a 3D perceptual similarity space. Such a visualisation of the Fb1 image selection for the term, 'embarrassment, shame' is shown in figure 7. The larger images represent repeat selections. The small grey dots represent the positions of the remainder of the 500 images in the browser. The greater the distance between two images in the view, the more dissimilar the original human image sorters judged them to be.

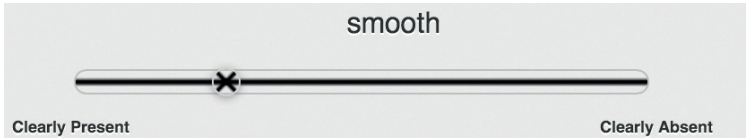
The image selections shown in figure 7 were divided into 10 clusters. One of the clusters is shown in figure 8. (The small red squares represent the rest the selections for "embarrassment, shame" in the remaining 9 clusters.)

The image nearest the centre of the cluster was chosen to represent the cluster. The summary collage of 10 perceptually representative images is shown in figure 9. The size is determined by the cluster population. The position in, 2D, on the collage is based on the image's position in the 3D MDS view.

During phase 2 of the experiment 40 visual stimuli were shown to naïve observers. The stimuli consisted of the output from Fb1 i.e. 20 image selections (sets of 60 images), one set for each of the 20 intended meanings, and the 20 corresponding summary collages. For each stimulus an observer used a screen of 20 visual analogue scale (VAS) (Hofmans and Theuns 2008) items to indicate their judgement of the degree to which the meaning of each term was present in the stimulus, see figures 10-11.

10

11



Each observer was shown 20 of the stimuli. Each observation consisted of 20 scores corresponding to the distance along the 20 term sliders where they placed their mark, see Table 1. In Table 1 the score in the slider for *w7* is 316. *w7* is 'enjoyment, pleasure'. 316 is the highest score (or first ranked) amongst all the term slider scores. If the stimulus being scored was one representing 'enjoyment, pleasure' then this would be an occurrence of 1st (or 1st equal) rank for the intended meaning of that stimulus. The maximum score on an item was set based on the length in pixels of the line on the VAS item as displayed on the iPad (Reips and Funke 2008).

The frequency with which observers allocated 1st rank to a stimulus' intended meaning *(f-1st)* was tallied up. It was normalised by dividing by the number of observers who were presented with that stimulus. The frequency of this occurring if the readings were to have been set at random was established by using a sampling technique. This randomly simulated 500 studies of 1000 readings each. (In our actual study the final number of observations was 20 for each of the 60 observers, totalling 1200). The results are shown in figure 12, along with random frequency level. The chart shows the *f-1st* figure for each of the 40 stimuli in the study. For the solid-set stimulus its score of 0.57 represents 17 out of the 30 observers who were shown that stimulus scoring the meaning, "solid", the top score among all 20 terms.

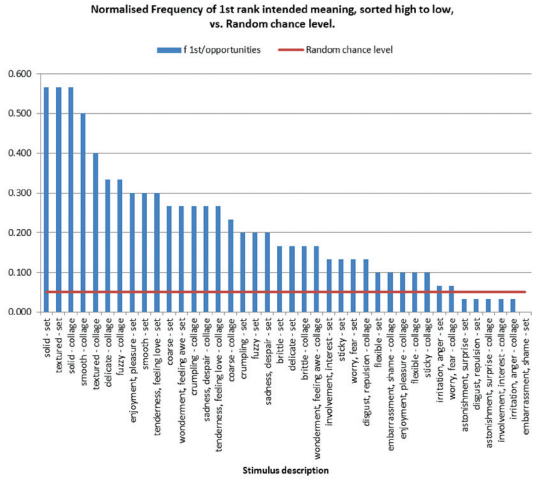
The frequency with which observers gave a stimulus top score for its intended meaning is well above the random level. The image sets chosen from our abstract image browser do convey meaning to the observers in the fb2 experiment. However the differences in the frequency figures show that the effectiveness does vary.

Figure 13 illustrates how often overall the stimuli were scored top or top equal for their intended meaning. In 250 observations out of 1200 the intended meaning was ranked 1st or 1st equal, representing 20.1% of observations. This indicates that there is a good deal of ambiguity of meaning in the visual stimuli. Notwithstanding this level of ambiguity the data do show the relative effectiveness of the selections from the abstract image browser in conveying descriptive and emotional terms, and the relative effectiveness of the image selections and the summary collages made to represent them.

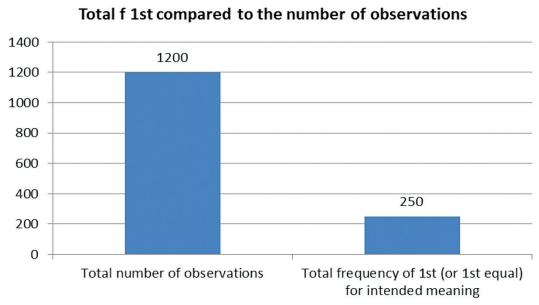
The normalised f-1st for stimuli derived from images chosen in stage 1 to represent descriptive terms, was compared to that for emotional terms. Figure 14 shows the mean f-1st for the 20 descriptive stimuli (10 sets and 10 collages) compared with that for the 20 stimuli which represented

w0	w1	w2	w3	w4	w5	w5	w7	w8	w9	w10	w11	w12	w13	w14	w15	w16	w17	w18	w19
303	211	144	70	290	6	16	316	0	54	220	19	51	191	258	13	287	128	309	10

Table 1

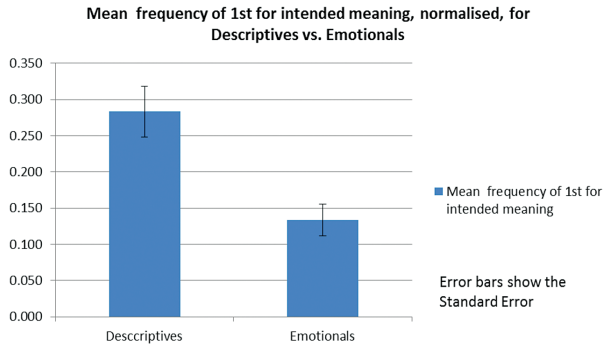


12

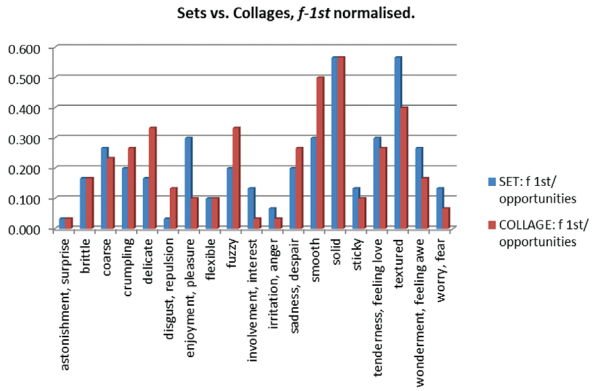


13

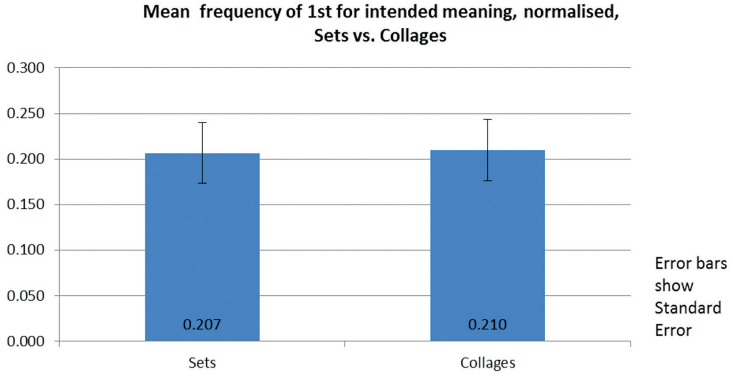
14



15



16



emotional terms. An independent t-test was applied to these two sets of data. On average, the normalised frequency with which stimuli representing descriptive terms had their intended meaning assigned a score ranking first or first equal, was significantly greater than for emotional terms.

We conclude that image selections made from our abstract image browser are more effective at conveying the meaning of descriptive terms than for communicating emotive terms.

The performance of the image sets compared to the corresponding summary collages did vary to some degree on a term by term basis, see figure 15.

However, when the two distributions are compared using a Pearson correlation the calculated correlation coefficient of 0.77 provides evidence that the two are strongly correlated (Field 2009). The mean f-1st across the 20 terms for sets was compared with that for collages, see figure 16.

A repeated measures t-test was applied to the two means. On average, the frequency with which the image set for a given term had its intended meaning assigned a score ranking first or first equal, was not significantly different to that of the summary collage for that term.

CONCLUSION

HeadCrowd has so far produced an intuitive abstract image browser which allows the handling of large amounts of images on handheld, touchscreen enabled devices. The project has collected and preliminarily analysed data on visual selection preferences of distinct 'crowds' when faced with distinct selection criteria.

We conclude that the summary collages constructed using our method

are equally as effective at conveying their intended meaning as are the full image sets on which they are based. This is an important finding as it means that if this communication system was used within a web site, thousands of images selected from the browser by crowds could automatically be represented by summary collages to deliver feedback of a manageable volume. The implications for co-creation in the fashion industry are wide ranging: If ideas and preferences of large crowds can thus be captured in a visual language, participation in the creative process can be widened beyond the small number of creative and commercial fashion leaders to include those who currently merely follow and consume. It can create buy-in and ownership for an industry that is maligned by the throw away nature of its products.

The research has now moved into using the collated data to question and adjust the choice and number of images a crowd may choose from, as well as considering the potential of the chosen selection terms for creative communications in the design process. Alongside this analysis, a cohort of 90 fashion and interior design students at level 9 is currently devising and testing visual, verbal and multi-media vocabularies for their usefulness in forecasting, advertising and creation of fashion design.

BIBLIOGRAPHY

Barnett, P (2003) Letters from the Editors. In Textile: Journal of Cloth and Culture, Vol 1, Issue 1, pp1-7

Blank Label (no date) Customize your own dress shirt [website] www.blanklabel.com/fabric (accessed 07/05/2013)

Chandler, D. (2002). Semiotics : the basics (2nd ed.): Routledge

Crow, David (2006) Left to right-the cultural shift from words to pictures. AVA Publishing SA, Lausanne

Field, A. (2009). Discovering Statistics Using SPSS. London, Sage

Hanington, B (2003) Methods in the making: A perspective on human research in design. In: Design Issues 19(4), pp. 9-18

Hebecker, R and Ebbert, H (2011) Visual Surveys with purposeful games. In: Information Design Journal 19(3) pp. 259-271

Hofmans, J. and P. Theuns (2008). "On the linearity of predefined and self-anchoring Visual Analogue Scales." Br J Math Stat Psychol. 61(Pt 2): Br J Math Stat Psychol. 2008 Nov;2061(Pt 2002):2401-2013. Epub 2007 May 2007

Kalkreuter, B and Robb, D (2012): Visual feedback for design. In Nordic Textile Journal, Sustainability and Innovation in the Fashion Field. 2012 (submitted)

Lawson, B (2005) How designers think. 3rd ed. Routledge

Levine, Locke, Searls & Weinberger (1999) The Cluetrain Manifesto [online] www.cluetrain.com (accessed 26th November 2012)

Methven, T. S., Orzechowski, P. M., Chantler, M. J., Baurley, S., & Atkinson, D. (2011, November 15 – 17). A Comparison of Crowd-Sourcing vs. Traditional Techniques for Deriving Consumer Terms. Paper presented at Digital Engagement '11, Newcastle, UK

Reips, U. and F. Funke (2008). "Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator." Behav Res Methods 40(3): 699-704.

Sanders, L and Simons, G (2009) A social vision for value co-creation in design [online] http://www.maketools.com/articles-papers/Social_Vision_for_Value_CoCreation_in_Design.pdf (accessed 10th June 2012)

Scherer, K. R. (2005). What are emotions? And how can they be measured? Social science information, 44(4), 695-729

United Styles (2012) Stand out from the crowd, design your own fashion [website] http://www.unitedstyles.com/en_usd# (accessed 07/05/2013)

FIGURE CAPTIONS

Figure 1: Proposed feedback system showing flow of ideas and information during a co-design or feedback cycle.

Figure 2: Card Sort: the first 100 images, using the method devised by Halley (2011).

Figure 3: The similarity matrix from the card sort and MTurk augmentation quantifying the perceived similarity between each image to every other image in the final set of 500. images.

Figure 4: 3D multi-dimensional scaling (MDS) visualisation of image relations, with two clusters picked out.

Figure 5: iPad interface of 48 stacks, with images within the bottom right hand stack shown.

Figure 6: Textural and Emotional terms used in Feedback experiments.

Figure 7: 3D MDS visualisation of the images selected to represent "embarrassment, shame".

Figure 8: 3D MDS view of one of the ten clusters from the "embarrassment, shame" image selection.

Figure 9: Summary collage for "embarrassment, shame".

Figure 10: iPad-1 with 20 VAS items and iPad-2 with one of the stimuli (in this case an image set).

Figure 11: One of the VAS item inputs. The X is placed by a tap and can be dragged back and forth. An observer could view the question by tapping a button on the iPad.

Table 1: The readings from a single observer scoring one image stimulus for each of the 20 terms using the VAS items.

Figure 12: Chart showing the normalised f1st for each of the 40 stimuli along with the level that would be expected were the readings to be set randomly.

Figure 13: Total frequency of first rank for intended meaning compared to the total number of observations overall.

Figure 14: Comparison of the f-1st for descriptive stimuli compared to that for stimuli representing emotional terms.

Figure 15: f-1st for sets compared to their corresponding collages.

Figure 16: Mean f-1st for the 20 image set stimuli compared the 20 summary collage stimuli.