Perceived Realism and Eye Tracking Performance within a Virtual Shopping Environment

David Stone
Clemson University
das@clemson.edu

Bobby Congdon
Sonoco Institute of Packaging Design & Graphics, Clemson University
rcongdo@clemson.edu

ABSTRACT

Understanding how consumers observe and interact with packaging has become increasingly more accessible and pertinent through the advent of eye tracking technology. In order to accurately predict how a package is going to sell or appeal to the customer, the test subject must feel and act as if they are in a natural shopping environment. If the process were executed virtually, this would create a more economical, customizable and faster way to investigate packaging stimuli. Using 3D modeling and real time rendering software, we created a virtual copy of our CUshop consumer experience shopping lab to compare with the physical lab. In order to negate the potential for cognitive dissonance within the study, video walkthroughs were recorded of each and are timed identically to minimize noticeable differences. The user's gaze is then tracked as they watch a video path moving through the grocery aisle in each environment. Gaze data combined with a modified version of the Witmer-Singer survey to gauge immersion within the environment allowed us to compare realism and eye tracking performance. Results revealed significantly lower scores in survey questions regarding Involvement and Sensory Fidelity, as well as differences in Total Fixation Duration between prerecorded video walkthroughs and the physical shopping environment.

INTRODUCTION

The packaging design workflow is heavily influenced by what is appealing to the consumer. Aside from asking someone to explain what led them to choose that product; you are limited to the few methods available for investigating packaging stimuli. Eye tracking has become one of the leading methods to answer the question of which package designs are actually holding consumers visual attention. Studies show that 90% of people make a purchase decision after only examining the front of the packaging and without having the product in hand. [Clement 2007] This means that the product's visual graphics can offer cues that act as a selling point for the product without the consumer having to interact with the package. Eye tracking analysis allows us to see the consumers' intentions quantified visually.

CUshop is the consumer experience lab used to conduct industry supported and student run eye tracking studies at Clemson University. The lab is designed to look and feel as if one is walking in a grocery store through the use of aisles lined with shelving, frozen food displays, and wall décor. Although many major companies have implemented this methodology, most seek it out after the product has been designed and fabricated. [Pieters 1999] This can create major losses in both time and money if the product isn't successful after the study. An alternative to testing how these packages are perceived would be to do it completely digitally. Advancements in 3D modeling and rendering software as well as graphics hardware have enabled the creation of highly realistic virtual environments.

Programs like Rhinoceros 3D and Google SketchUp can be utilized to model, script and render an environment all within the same program. Modeling software integrated in to the graphic and dieline design process allow for a large number of possibilities for packaging visualization. The challenge is taking a realistic looking product and presenting it in an environment that doesn't convolute the ability of the consumer to analyze and decide as they would in a grocery store environment. We used a program called Esko Studio Store Visualizer to create the shopping environment. This program allows one to import a 3D Collada file in to a scene that has pre-rendered lighting, shadows and reflections to create a realistic looking product in context. Each object imported in to the scene can be adjusted in scale, color, reflection, gloss, physical properties etc. When all of these tools are combined, we are able to design an immersive environment to conduct eye tracking studies in comparison to a real world shopping environment. This study will continue to explore whether eye tracking studies in computer generated environments will be a viable substitution for packaging design evaluation currently conducted in real stores.

BACKGROUND

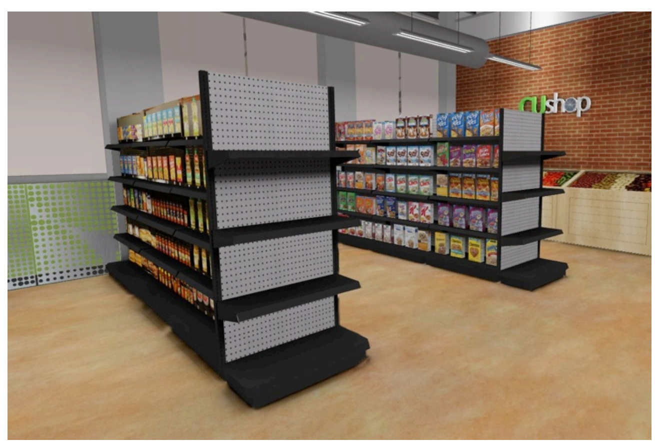
Understanding the connection between purchasing patterns and packaging design has been a topic of discussion for decades. James Pilditch termed packaging design as *the silent salesmen* for a company's branding. [Pilditch] Since majority of the purchase decision is made without actually touching the package, it can be said that "what you see is what you choose." [Clement 2006] So to gather an understanding of what sells, we must understand how the package is perceived. There have been several methodologies for the process in which people analyze products, but Ursula Hansen seems to break it down in to its simplest terms. She said that packaging has influence on buying behavior through three characteristics: communication, functionality and environment. [Hansen 1986] Communication focuses on the primary visuals of the package (graphics, color etc) that help initiate the first fixation on a product. The functionality pertains to how the consumer interacts with the package from shelf to actual usage. Lastly, the environment consists of shelf life to disposal of the product (recyclability, reusability etc). [Hansen 1986] Although these three functions of the package are not necessarily considered at once, they are accounted for at some point within the life of package. Majority of consumer studies are run through focus groups and interviews but could be much more effective if the data were able to be quantified through what and how long they looked at the target objective.

Quantifying the time and location of a participant's focus can be accessed using eye tracking tools in conjunction with a convincing environment. A shopping context is said to be one of the most important contributing factors to obtaining viable data from consumer packaging research. [Young 2004] Meaning that within the shopping environment, the package that stands out most will typically be chosen first. Although not always true, it can be said that packaging that uses distinct features like shape, orientation, color and size are more likely to attract and influence peoples purchasing decisions. [Clement 2006]

A shopping context is not only associated with content but also with the environment in which the study takes place. Tonkin et al. study on shopping with projected shelves vs. a physical environment showed that consumers completed their search task significantly faster when immersed in a real world setting than in a virtual environment. [Tonkin 2011] Tonkin indicated that this is due to the fact that "the number of fixations generally coincides with time taken to complete visual search." Meaning that the participant possibly took more time on the projected shelf due to reasons like the fidelity of the projected image or higher fixation rate due to the amount of time required in the search for the target product.

Visual realism can be one of the major contributing factors to the success of data collection within a virtual environment. In Witmer and Singers' study about presence in a virtual environment, they said that fully immersed observers feel as if they are actually interacting directly, not indirectly with the given environment. [Witmer & Singer 1998] Realism can be influenced by continuity and consistency that is objectified through previous real world interactions within shopping environments. In order to achieve this validity a virtual environment must feel and look like a real environment. An ideal virtual environment would incorporate the three varieties of realism as mentioned by Ferwerda: Physical realism (visual stimulation in the scene) photo realism (visual response in the scene) and functional realism (visual information in the scene). [Ferwerda 2003] However, creating and rendering a space that incorporates all of these types of realism is unlikely due to the number of variables involved. There will always be a disconnection between the monitor and the user when moving in virtual world since they can't actually be inside of the space. Our study of virtual shopping intends to explore these issues, focusing on creating an immersive virtual environment able to capture consumer data equal or comparable to that which is recorded in real world situations.

(Figure 1: Photo of CUshop Consumer Experience Lab at Clemson University.



(Figure 2:: CUshop Lab modeled with Esko Studio Store Visualizer.)

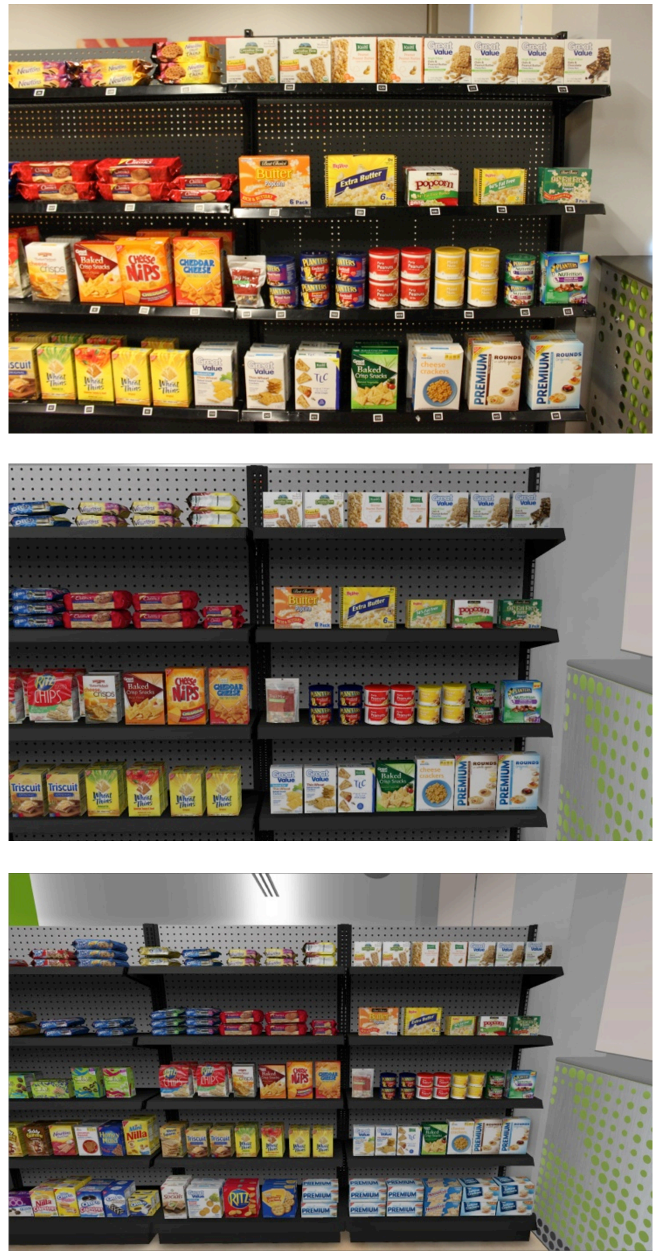
METHODOLOGY

Measuring the virtual CUshop in comparison to the real CUshop was executed by determining the person's performance in the given environment using eye tracking data and the perceived realism of the shopping experience.

Stimulus

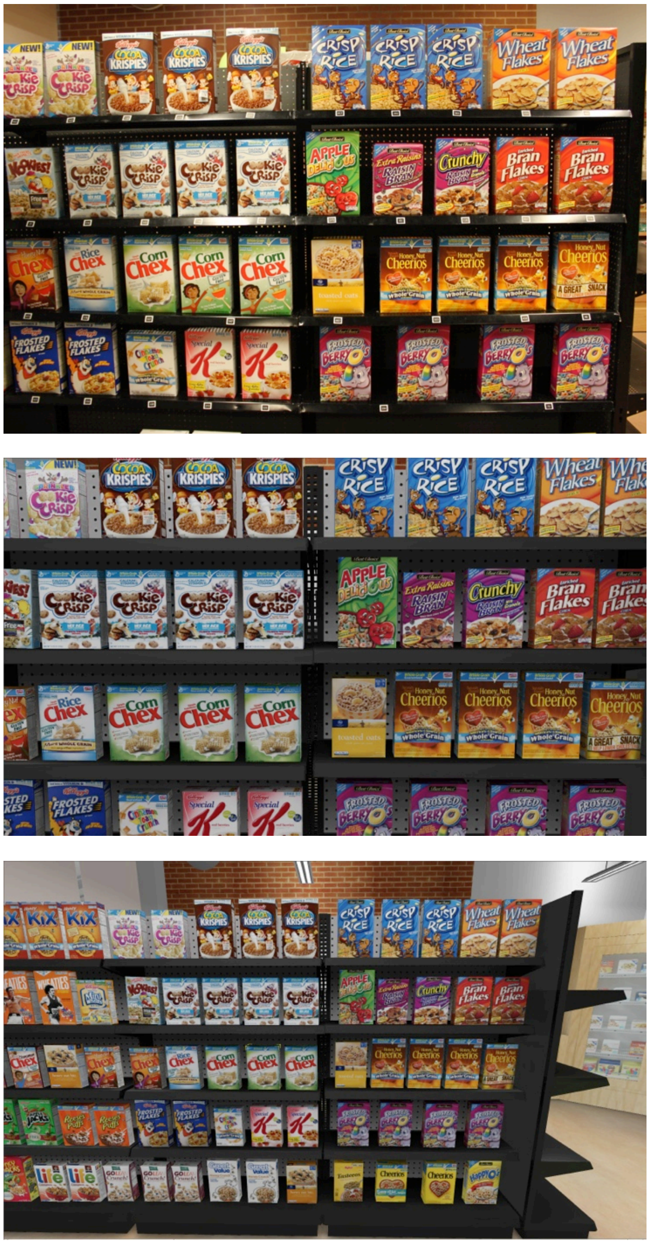
The CUshop consumer lab (Figure 1) consists of three aisles, each equipped with shelving units that are individually 4’ in width and 1’6” in depth and 7’ in height. The shelving units are dark gray with lighter gray pegboards to create a backdrop for the shelves. Each aisle length is 12’ with a 7’ walkway in between sets of 4 shelves. The walls parallel to the aisles are lined with fruit shelves and three mock-frozen shelves for displaying frozen food items. Flourescent lighting is used to simulate the lighting you would experience in a grocery store. Lastly the floor of the lab is a semi-glossy brown marble texture that provides slight reflections of both the shelves and products displayed. To capture the video walkthrough of CUshop lab for the experiment, a Canon Rebel T1i was mounted to a tripod with wheels and pushed along a predetermined path throughout the lab. A pilot study with three graduate students found that the mean time spent in front of each shelf searching for the product was approximately eight seconds. Therefore, the camera path paused at the two target shelves for eight seconds each to allow adequate search time to locate the product. This video stimulus will be referred to as *VCU (Video of CUshop).*

The Virtual CUshop (Figure 2) was measured and replicated to be exactly like the consumer lab. This required creating custom shelves to match the ones in use, in addition to modeling the details/objects (walls, air ducts, doors, windows, furniture, etc.) within the consumer lab. Using Rhinoceros 3D architectural modeling software, we built the structural environment and imported Collada files as objects in to Esko Studio Store Visualizer. When importing in to Store Visualizer, each model had a custom UV map that generated in to a texture map .png that could be altered in Photoshop to create realistic textures (i.e. marble floor, wood paneling, wall colors, etc.), shadows and reflections. Once the models were placed in Store Visualizer, each was precisely positioned to match the physical lab. After textures were applied, we added to the aesthetics of the room (light switches, paintings, logos, exit signs, base boards etc.). After creating the environment we established two fields of view with the scene camera at 60mm (*VR60*) and 90mm (*VR90*). This compensated for the notion that peripheral vision is used to widen the search beyond a cameras limited field of view. [Tan et al. 2006] A video sequence was generated using camera position presets placed in the scene and timed to sync up with the video recorded in the CUshop consumer lab. When played, the sequence moved and rotated around to simulate the participant walking through the virtual grocery store. The videos were exported as high resolution .avi files at a frame rate of 24.



(Figure 3: 94% Fat Free Popcorn target shelves in each of three stimuli. Top: VCU & Physical, Middle: VR60, and Bottom: VR90.)

Stimuli for the experiment were based on comparing the effectiveness of Raisin Bran cereal and 94% fat free popcorn in comparison to its competitors. Each product was scanned using an Epson Perfection V30 scanner at 300 dpi and imported in to Adobe Illustrator CS5. Once in Illustrator, a custom die line was generated using Esko Studio plugins and graphics were applied. The 3D model was exported as a Collada archive file (.zae), then imported in to Esko Studio Store Visualizer. Shelves were populated with store and name brands of cereal, cookies, popcorn, and crackers to match the physical CUshop.



(Figure 4: Raisin Bran Cereal target shelves in each of three stimuli. Top: VCU & Physical, Middle: VR60, and Bottom: VR90.)

Apparatus

Two methods were used for capturing eye tracking data within the physical and virtual CUshop.

<u>*Physical CUshop*</u>: Eye movements for the physical CUshop were captured using Tobii eye tracking glasses (Figure 5). The glasses feature a monocular lens with a recording rate of 30 Hz.[Tobii] They are also paired with IR markers to specify which product was being viewed on the shelf via a plane created in space called an AOA (Area of Analysis). The AOA gives feedback on a specific location within the testing environment to show fixation on the testing product. The eye tracking glasses are connected to a device called the *Recording Assistant* which stores both the calibration and gaze data from the glasses. In addition to the data, the device also records video to be referenced when analyzing data from the experiment. Calibration for the glasses require an IR marker to be placed and moved at 9 different points on a vertical plane in front of the test subject.

(Figure 5: Tobii eye tracking glasses and recording assistant device)

<u>*Virtual CUshop:*</u> Data collection for the virtual CUshop was captured by a Tobii T60XL eye tracking monitor. The screen featured a high resolution 24-inch TFT wide screen monitor (Figure 6) to allow for wider screen gaze angles and large head movements. All precision measurements are recorded at a 60 Hz sampling rate and distance of 65cm with a processing latency of less than 17 ms. [Tobii] The monitor has cameras within the frame of the screen that record eye movement of what is being displayed on the screen. When calibrating, the user is asked to watch a red dot as it moves around the screen to 9 different points while refraining from moving their head as little as possible.

(Figure 6: Tobii T60XL eye tracking monitor)

Experimental Design

The experiment was either conducted in the physical or virtual CUshop. Each participant could only participate in one study to negate the possibility of already knowing where the product is placed. The target products were set off center of each screen in order to avoid center fixation from occurring within the walkthrough videos. [Tonkin 2011] The participants were told to shop for a Raisin Bran Cereal and 94% fat free popcorn box within the aisle.

Participants

The study consisted of 126 participants ranging from ages 20-65. Due to various errors in recording, 13 of the total participants were eliminated since the data was unusable. The location of the experiment was at PackExpo, where we were able to test a large demographic of both ages and genders. The test ran over a period of 3 days and participants were randomly assigned to watching one of the 3 videos recorded (*VCU, VR60 and VR90*) or "shopping" in the physical store. Seventeen participants watched VCU, 20 watched VR60, 22 watched VR90, and 54 walked through the physical store.

Procedure

The physical CUshop experiment began by calibrating the participant using the 3 x 3 grid for calibrating the eye tracking glasses. Once calibrated, they were handed a shopping list consisting of target products including Raisin Bran cereal and 94% fat free popcorn. After entering the store, they search the aisles looking for the target items and write down which they would have actually purchased. Each target item was paired with a reference number attached on the shelf below. After shopping, they exited the store, ended the eye tracking recording and removed the glasses. The researcher then led them to a survey computer to answer questions from a modified Witmer Singer survey in conjunction with questions about demographics, eye conditions and age range.

Immersion in the virtual CUshop began by bringing the participant in to a closed room with a Tobii monitor. After asking them to take a seat, the monitor was adjusted to their height to ensure the eye tracking cameras could locate their gaze. The researcher then prompted them to follow the red dot on the screen with their eyes as closely as possible for calibration. Once completed, they were asked a few questions pertaining to demographics and whether or not they had participated in the physical CUshop experiment. Once finished with the survey, the participant is prompted with a search task followed by a video (*VCU, VR60 or VR90*) walkthrough of either the virtual or physical walkthrough. During the video the camera moves in to the virtual/physical environment and stops for 8 seconds at each shelving environment to search for the target product. Lastly, the participant is asked fourteen questions from a modified Witmer-Singer survey [Tonkin 2011] that was tailored to comparing the environments in our study. Questions were divided into four categories, randomized, but presented in the same order to each participant. The survey was administered within Tobii Studio for the video-based stimuli, and with an online survey taken on a laptop computer for the physical store.

RESULTS

Each video-based stimuli was evaluated versus the physical store by means of modified Witmer-Singer presence questionnaire scores and eye-tracking performance metrics of time to first fixation (TTFF) and total fixation duration (TFD).

Presence Questionnaire

The modified Witmer-Singer survey used consisted of fourteen questions each ranked on a 7-point Likert scale. Questions were categorized as Involvement, Immersion, Sensory Fidelity or Interface Quality. Four of the questions were negative in nature, so scores were transformed to be consistent with the other questions, in which a high score indicated a positive response. An ANOVA for mean total score (all 14 questions) showed no significance. However, an ANOVA for the means of each of the question

Table 1: Mean responses for the modified Witmer-Singer presence questionnaire using a 7-point Likert scale with 7 indicating highest level of agreement and 1 indicating lowest level of agreement. Statements 5, 6, 12 and 13 are negative in nature, and values were transformed when calculating Group Means for score consistency.

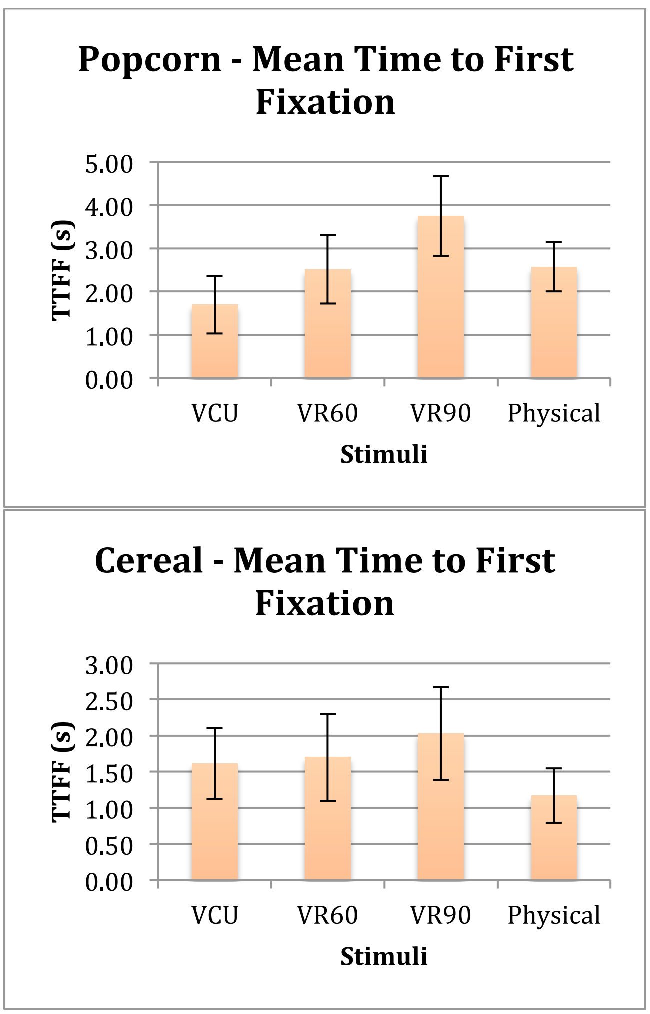
#	Question	VCU	VR60	VR90	Physical
	Involvement				
1	*My interactions with the shelving environment seemed natural.*	4.41	4.65	5.18	4.98
3	*The visual aspects of the environment involved me.*	5.06	5.25	5.32	5.63
8	*I was able to completely survey or search the environment using vision.*	5.00	5.05	5.64	5.80
11	*I felt involved in the search task.*	4.76	4.60	4.77	5.85
	Group Means (Mean of Means)	4.81	4.89	5.23	5.56
	Immersion				
2	*All my senses were completely engaged.*	4.59	4.15	4.82	5.20
4	*I was completely aware of events occurring in the real world around me.*	4.65	4.40	4.50	4.69
6	**The information coming from my visual sense felt inconsistent or disconnected.*	*4.06*	*3.55*	*3.00*	*3.43*
12	**I was distracted by display devices.*	*3.12*	*2.75*	*2.36*	*3.22*
	Group Means (Mean of Means – Adjusted for Negative Statements)	4.51	4.56	4.99	4.81
	Sensory Fidelity				
7	*My experiences with the shelving system seemed consistent with my real-world experience.*	5.24	4.75	4.91	5.30
9	*I felt that I was able to examine objects closely.*	3.76	4.30	3.86	5.41
10	*I felt that I was able to examine objects from multiple viewpoints.*	3.88	3.85	3.86	5.22
	Group Means (Mean of Means)	4.29	4.30	4.21	5.31
	Interface Quality				
5	**I was completely aware of any display and control devices.*	*4.47*	*4.05*	*4.32*	*4.67*
13	**Visual display quality interfered or distracted me from completing my task.*	*3.24*	*3.00*	*3.55*	*2.91*
14	*I was able to concentrate on the search task and not on the devices used to perform the task.*	5.29	5.20	4.36	
	Group Means (Mean of Means – Adjusted for Negative Statements)	4.53	4.72	4.50	4.21

subcategories revealed significance for Involvement ($p<0.02$) and Sensory Fidelity ($p<0.01$), but not for Immersion or Interface Quality.

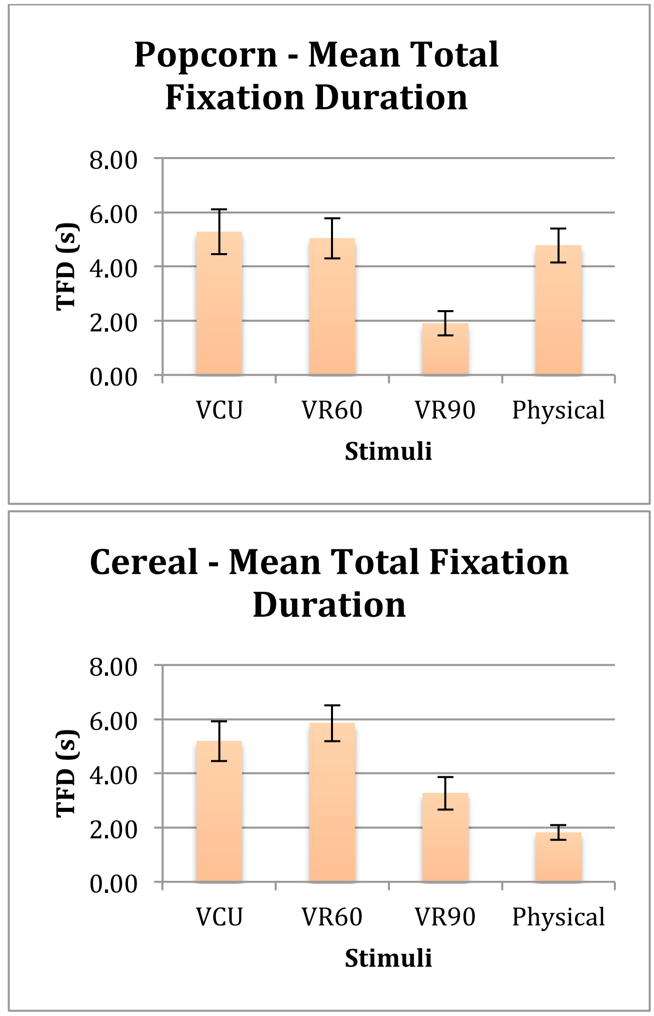
Eye-Tracking Metrics

Eye-tracking metrics were calculated in Tobii Studio 3.1 using the Tobii fixation filter, and calculating gaze point using the Average method for stimuli recorded on the Tobii XL60 monitor (VCU, VR60 and VR90). Data was exported for each of the two search tasks (raisin cereal and 94% fat free popcorn) for all three video stimuli and the physical store. An ANOVA showed no significance for TTFF for either search task, but significance for TFD on the cereal ($p<0.01$) and the popcorn ($p<0.01$). For the popcorn, the VR90 had the lowest TFD. For cereal, the physical store had the lowest TFD.

Given the fixed amount of time in each of the video stimuli, it is worth noting the number of participants for each that did not locate the popcorn packages (VCU: 5.8%, VR60: 10%, and VR90: 18.2%). However, this data proved statistically insignificant.



(Figure 7: Mean Time to First Fixation for each target item across the four stimuli environments.)



(Figure 8: Mean Total Fixation Duration for each target item across the four stimuli environments.)

DISCUSSION

The lack of significance in mean presence survey scores was surprising, although the data showed a trend. After survey subcategory analysis, the significance found in the Immersion and Sensory Fidelity question sets was more in line with predicted outcomes. The questions from these subcategories are predominately based on free movement and control. Since all three video-based stimuli were prerecorded walkthroughs with a set point of view and timings in front of the shelves, we expected lower perceived immersion scores. This confirms the notion that being able to move around at will in an environment is critical to the perceived realism of the participant.

Participants not having navigational or temporal control seemed to have an affect on the number of participants failing to locate the target product on the shelf, although that affect proved insignificant. We believe this finding is noteworthy nonetheless, as any number of participants being unable to find the product on the shelf points to an insufficient method of conducting packaging assessment. It is difficult to determine why some participants were unable to locate the product. Poor screen resolution, not enough time in front of the shelf, unclear search objectives, or technical issues with the eye-tracker may have contributing factors.

The walkthrough of the virtual environment with the 90mm camera lens setting (VR90) was intended to provide participants with a wider field of view and therefore locate products faster with periphery mapping of the shelf. The TTFF data showed that despite mean times being longer for VR90, it was an insignificant difference. This may be due to the reduced size of the target packages when displayed on the monitor. It will be interesting to repeat the study with adjustments made to the display size of each video in order to correct the apparent size of each package on screen to be identical. Previous research [Tan et al. 2006] indicates that the extra information displayed in periphery should improve search performance.

The recorded video walkthrough of the physical store (VCU) is admittedly not a practical method of conducting eye-tracking research, but provided a means with which to judge the accuracy of the constructed virtual store. Eye-tracking performance and presence questionnaire scores were found to be of insignificant difference between the two, indicating sufficient accuracy. This is not to say that the recorded walkthrough of a virtual store is a sufficient solution for packaging evaluation. However, using the metrics we evaluated, the level of detail available in the software is on par with a high-resolution video capture of a physical store shelf.

STUDY LIMITATIONS

As noted in the previous Discussion section, the size of the target packages in pixels when displayed on the monitored varied between video stimuli, likely skewing eye-tracking metrics and ability to locate products on the shelf in a timely manner.

While methods of calculating TTFF across video stimuli on the Tobii monitor were consistent, the data captured in the physical environment with the Tobii glasses is more difficult to determine start times for calculating TTFF.

Participants in the physical environment were instructed to mark on a shopping list the identification number of the product they were selecting, whereas participants on video stimuli were not given any instructions for what to do after locating target products. This may have led to excessive fixations on target products after the search task was complete and the participant was waiting for the video to continue. Several participants during the video stimuli commented on not knowing what to do once they had located the target product. Others tried to either touch the screen or click on the product with the mouse.

The Esko Store Visualizer software accommodates navigation through the virtual store, but plans to conduct eye tracking while recording the screen were cancelled due to incompatibilities between Tobii Studio and Store Visualizer. Developers from both companies are now aware of the issue.

CONCLUSIONS

We have analyzed various methods of conducting monitor-based search tasks for packaging on a shelf versus the same search task in a physical lab environment. An ideal virtual environment would score similarly on qualitative measures of participant presence, and would gather similar data on important eye-tracking metrics such as TTFF and TFD. Although mean scores for presence were not significantly different, two of the four subcategories in our questionnaire pointed to shortcomings in the video stimuli, presumably related to lack of navigational control. Total Fixation Duration was the only eye-tracking metric examined that showed significance, although results between target products varied in which stimuli they performed poorest. We can conclude that none of the three video-based stimuli presented will yield accurate results for packaging evaluation. However, results suggest that the level of detail and rendering available in the store visualization software represent packaging on a store shelf with a sufficient degree of accuracy.

FUTURE WORK

The scope of this study was limited by available equipment for displaying and tracking using a virtual environment. Future studies will make use of new technologies as they become available to us and explore the aspects of a virtual environment that affect the perception of realism and the eye-tracking performance of participants. Variables affecting viability may include stereoscopic effects, resolution, navigational controls, display size relative to participant, and more. Each variable should be examined and balanced with cost while building an acceptable virtual test environment. Once software incompatibilities are resolved and free navigation is possible, experiments will be conducted to compare with stimuli presented in this paper.

Accommodating a means of product selection either within the software or on paper, in conjunction with a self-paced exploration of the virtual environment, may lead to more accurate results and higher perceived realism.

Furthermore, methods of evaluating a virtual environment for realism and performance should receive more consideration. Physiological measurements may prove useful in comparing virtual to real correlation.

Technical limitations of a virtual environment for quantifying consumer response to packaging design should also be explored. Specialty printing and finishing techniques (i.e. matte coatings) may prove difficult to discern differences.

ACKNOWLEDGEMENTS

We would like to thank Esko for their ongoing support and software donations, Tobii for their support and use of hardware, and PMMI for the invitation and support for our PackExpo 2013 booth.

REFERENCES

Clement, J. (2006), “Packaging design taxonomy”, *Design to business, D2B the 1st international design management symposium*, Shanghai, Vol. **1***,* No. 1 pp. 150-158.

Clement, J. Visual influence on in-store buying decisions: an eye-track experiment on the visual influence of packaging design. *Journal of Marketing Management*, 23(9):917 — 928, 2007.

Ferwerda, J. A. 2003. Three Varieties of Realism in Computer Graphics. In Human Vision and Electronic Imaging. SPIE, Bellingham, WA, 290–297.

Hansen, U. (1986), “Verpackung und konsumentenverhalten (packaging and consumer attitudes)”, *Marketing,* Vol. **8**, No. 1, pp. 5-12.

Pieters, R. and Warlop, L. (1999), “Visual attention during brand choice: The impact of time pressure and task motivation”, *International Journal of Research in Marketing,* Vol. **16**, No. 1, pp. 1-16.

Pilditch, J. (1973), *The silent salesman* (2nd ed.), London: Business Books Limited.

Tan, D. S., Gergle, D., Scupelli, P. G., and Pausch, R. 2006. Physically large displays improve performance on spatial tasks. In ACM Transactions on Computer-Human Interaction. Vol. 13. ACM, New York, NY, 71–99.

Tonkin, C., Ouzts, A. D., and Duchowski, A. T. 2011. Eye Tracking Within the Packaging Design Workflow: Interaction with Physical and Virtual Shelves. In Proceedings of Novel Gaze-Controlled Applications. ACM Press, Karlskrona,
Sweden.

Young, S. (2004), “Breaking down the barriers to packaging innovation”, *Design Management Review,* Vol. **15**, No. 1, pp. 68-73.

Witmer*,* B. G. and Singer, M. J. 1998*. Measuring Presence in Virtual Environments: A Presence Questionnaire.* Presence 7, 3 (June), 225–240