Perceptual Importance of Lighting Phenomena in Rendering of Animated Water

MICAH BOJRAB, MICHEL ABDUL-MASSIH, and BEDRICH BENES, Purdue University

Recent years have seen increasing research in perceptually-driven reductions in the costs of realistically rendered imagery. Water is complex and recognizable, and continues to be in the forefront of research. However, the contribution of individual lighting phenomena to the perceived realism of virtual water has not been addressed. All these phenomena have costs associated with their rendering, but does the visual benefit outweigh these costs? This study investigates the human perception of various illumination components found in water-rich virtual environments. The investigation uses a traditional psychophysical analysis to examine viewer perception of these lighting phenomena as they relate to the rendering cost, and ultimately reveals common trends in perceptual value. Five different scenes with a wide range of water and lighting dynamics were tested for perceptual value by one hundred participants. Our results provide an importance comparison for lighting phenomena in the rendering of water, and cost reductions can be made with little or no effect on the perceived quality of the imagery if viewed in a scenario similar to our testing.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation—*Viewing algorithms*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Color, shading, shadowing, and texture*

General Terms: Design, Experimentation, Performance

Additional Key Words and Phrases: Perceptual importance, water animation, shading effects, illumination

ACM Reference Format:

Bojrab, M., Abdul-Massih, M., and Benes, B. 2013. Perceptual Importance of lighting phenomena in rendering of animated water. ACM Trans. Appl. Percept. 10, 1, Article 2 (February 2013), 18 pages.
DOI = 10.1145/2422105.2422107 http://doi.acm.org/10.1145/2422105.2422107

1. INTRODUCTION

Water is a common, recognizable substance; its rendering is complicated, and usually time-consuming. Humans perceive water through its interaction with light and the environment surrounding it. The most important lighting phenomena when water is rendered can be roughly categorized into specular reflection, refraction, caustics, and shadow. These phenomena, coupled with water's complex surface, give it a unique and complicated appearance.

An important part of previous work on fluid simulation and rendering has been focusing on physically-accurate, photorealistic visualizations such as [Jensen 2001; Donner et al. 2009]. Physically-accurate fluid rendering presents one of the most computationally challenging tasks in computer

Author's address: B. Benes, Department of Computer Graphics Technology, Knoy Hall of Technology, Room 313 401. N.Grant Street, West Lafayette, IN, 47907-2021; email: bbenes@purdue.edu.

DOI 10.1145/2422105.2422107 http://doi.acm.org/10.1145/2422105.2422107

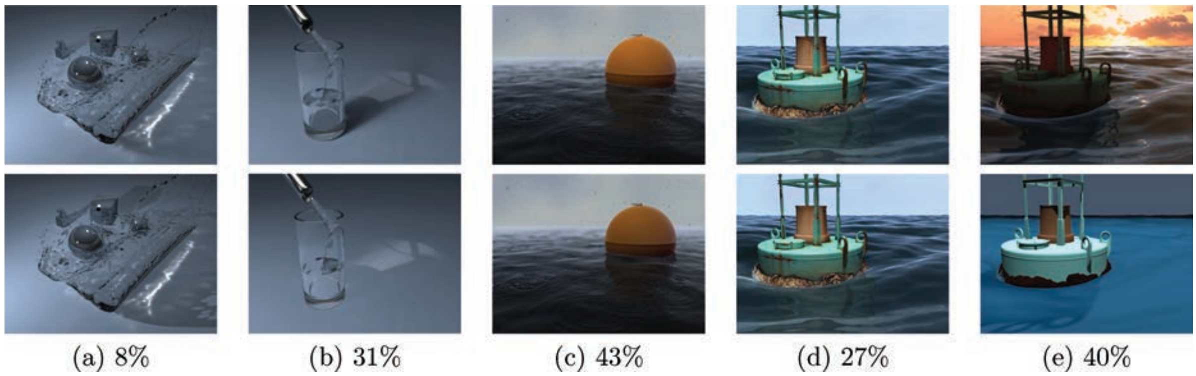

Fig. 1. This figure shows pairs of images from the five testing scenes used in our psychophysical analysis. The upper row shows the gold standard rendering, which contains the natural lighting phenomena of the tested variables. The lower row displays a matching image from the same scene where a random illumination component is either missing or substituted: (a) hard shadows substituted with soft shadows; (b) no shadows present; (c) no caustics present; (d) no shadows present; (e) diffuse color substituted with refraction. The percentage of savings in rendering is displayed below the images.

graphics. On the other hand, recent years have seen research focused on the perceptually driven rendering techniques that can reduce the rendering costs while maintaining a relatively high visual quality. Perceptually driven scene simplifications in terms of colors for tone mapping (e.g., [Irawan et al. 2005; Masia et al. 2009]), variance of objects in the scene [McDonnell et al. 2009], and polygonal count [Ramanarayanan et al. 2008] have been found to reduce the effect on visual degradation in the perceived imagery. A possible avenue is using the characteristics of the human visual system (HVS) to alter specific rendering techniques [Cater et al. 2003; Chalmers et al. 2006; Dumont et al. 2003]. Various methods have been proposed using what is known about the HVS, and techniques that allow reducing rendering times while maintaining high visual quality have been discovered [Debattista et al. 2005; Stokes et al. 2004]. Our work builds on this previous research and extends it by using a traditional psychophysical analysis coupled with a rendering cost analysis to provide an importance comparison of the lighting phenomena in an audience's perception of animated water. The assumption of our work is that not all lighting phenomena in water-rich virtual environments are perceived with the same importance. Water is rarely the main focus of the scene, and it typically resides in the background, which further complicates illumination importance. The combination of having high rendering costs, low visual priority, and ubiquity makes water a great candidate for the rendering time reduction by using perceptually driven techniques.

We have created five scenes showing various, but common, water-rich scenarios with the most important lighting phenomena. Our scenes encompass a wide variety of cases, ranging from shallow to deep water with low- and high-motion dynamics. The scenes also convey low- and high-dynamic ranges of lighting. We have tested the visual perception of the scenes on one hundred human subjects. They have ranked the lighting phenomena based on their perceived contribution to the visual quality of the scenes. Visual quality is based on the realism conveyed by a scene as perceived by the observers. Though, some scenes rendered in our study were not photorealistic or physically rendered, the visual quality of a scene was a factor of how the viewer perceived its realism and how close that reaction would be to a real scene.

Our results suggest that for most scenes used in our testing, as much as 40% of rendering costs can be reduced without significant degradation in perceived quality, compared to rendering of the same scene with all the lighting effects present. Moreover, results show a significant difference between shallow and deepwater scenes. The main contributions of our work include the following:

(1) providing a comparison of importance in lighting phenomena specific to water;

(2) using a psychophysical analysis to obtain pure subjective quality of lighting elements; and

(3) combining the rendering cost with the visual quality to obtain relative importance.

The article continues with the description of the previous work. Sections 3 and 4 describe the testing variables and methodology, respectively, and Section 5 discusses results of that testing. The article concludes with Section 6, which summarizes the article and discusses avenues for future work.

2. PREVIOUS WORK

The computer graphics community has dedicated much effort and research into the understanding and proper representation of fluids for more than three decades. Recent research has shown that fluids continue to be an important topic, and that better, faster, and more physically accurate algorithms for their simulation and rendering are developed each year. A detailed description of the previous work in this area is out of the scope of this article and we just highlight the most important trends.

Fluids can be rendered in various ways, one of which includes nonphotorealistic rendering. Other techniques, typically used in real-time rendering and gaming, include fast and usually approximated approaches. Both of the above-mentioned classes of algorithms are out of the scope of this article. The third class of techniques, and the focus of this study, is the ever-increasing trend toward photorealism. Enright et al. [2002] described a new photorealistic system for animating and rendering water with a special focus on the free level of water. Some effort has been dedicated to the rendering of wet materials [Donner and Jensen 2005]. An alternative method for a photorealistic rendering of refractive and reflective surfaces was photon mapping, which was proposed by Jensen [2001]. Improvements to this system have been focused on reducing rendering time while maintaining quality through progressive multipass methods [Dmitriev et al. 2002; Hachisuka et al. 2008; Hachisuka and Jensen 2009]. Recent work has shown progress towards the rendering of dissolved materials in water by using BSSRDF [Donner et al. 2009].

Recent research has changed reduction techniques into a psychological study. The research is becoming more interdisciplinary, looking not only into the reductions, but also into how these reductions affect the subjective quality to humans. The human visual system (HVS) is an important topic in techniques that attempt to reduce rendering time based on human perception. Past studies have made these algorithms based on the various natural limitations of the HVS, namely, through inattentional and change blindness [Cater et al. 2002, 2003]; and selective rendering using saliency maps [Chalmers et al. 2006]. Nasr et al. [2006] applied perceptually driven rendering techniques to video games; Anson et al. [2006] investigated the perceptual importance to rendering of participating media; and Debattista et al. [2006] applied perception in parallel rendering using irradiance caching. McDonnell et al. [2008] and [2009] investigated the perception of crowd simulations and the variation in characters within. Ramanarayanan et al. [2008] explored the extent of perceptually driven geometry reduction for complex aggregates. Techniques exist for perception-guided tone mapping [Masia et al. 2009]; image differentiation determination for high dynamic range imagery [Irawan et al. 2005]; multiresolution color models [Tolhurst et al. 2005]; and displays [Didyk et al. 2010]. Alternative methods based on visual quality identify an individual's percept, or emotional response, to stimuli. These studies generally conduct a psychophysical analysis of imagery to gain information on contrast and quality. Researchers have used contrast data to find the thresholds of change and to introduce reductions below this perceivable amount. McNamara et al. experimented with the perception of real and rendered images in McNamara [2005]. Fast and approximated shadow rendering was addressed in Sattler et al. [2005]; and Kozlowski and Kautz [2007] investigated glossy reflections in virtual scenes. Other researchers have used quality data in the application of selective component-based rendering to direct resources to the lighting elements that contribute the most to visual quality [Stokes et al. 2004; Debattista et al. 2005]. Furthermore, the perceptual influence of approximate visibility in indirect illumination was addressed in Yu et al. [2009]. Motion in a scene may affect perception, and this phenomenon was recently addressed by Hasic et al. [2007], Hasic and Chalmers [2009].

Our study builds on the psychophysical analysis as described in Stokes et al. [2004] and Yu et al. [2009] to determine the subjective quality of the lighting phenomena in water-rich scenes. We extended this by considering costs associated with different lighting phenomena. Moreover, water is usually not the central theme of the scene, so the visual quality of lighting can become diminished even more. A new aspect not offered by previous research is that we compare not only static images in our psychophysical analysis, but also water animations.

3. TESTING PARAMETERS

Here we describe the testing variables and scenes used in our psychophysical analysis. The objective was to individualize the lighting components and create general scenes that convey a wide variety of practical water-rendering cases. The key to our approach is a component-based method of illumination. Our work builds on the same assumption as the previous work of Debattista et al. [2005]; Stokes et al. [2004]; and Yu et al. [2009], which divided the illumination into components that can then be tested for visual quality by means of a psychophysical study. Our hypothesis is that not all illumination components are equally important to the realism of water rendering, and lighting elements can be substituted or subtracted without perceiving a significant visual difference.

The participants were intentionally uninformed of the rendering style, content of the scene, and subject matter. The results of the research could be used in digitally rendered animations displayed in a movie, where the audience will have much less time to strictly analyze the image. We have created five testing *scenes* with different water dynamics, each rendered eight times with different illuminations and material components. We will refer to these individually rendered scenes as *animations*. In all, we have created 40 animations, each lasting over two seconds.

We have intentionally decided to use a rendering of water and not a real video. It is virtually impossible to recreate the video exactly, so using a video of water against animations might create a potential problem of highlighting differences in salient features, such as the quality of the animation or some other subtle effects that could severely invalidate the results.

3.1 Illumination and Material Components

Elements tested in our study are divided into the two main categories: materials and lighting elements. Water material properties include diffuse transparency, diffuse grey, diffuse color (which are both typically used in previsualizations), and refraction (see Figure 2). Material variables were used to define the surface of the water, and no two materials were applied at the same time. The lighting elements include specular reflection, caustics, hard shadows, and soft shadows.

Scenes were modeled in Maya and rendered in Mental Ray eight times. Mental Ray is used in animation production, and it is one of the state-of-the art renderers providing all the functionality that we tested. Each scene contained different illumination and material settings. The gold standard is included in the psychophysical analysis, as the control and is assumed to be the most visually precise because it contains the most physically correct lighting and material elements, and was rendered with the most realistic options available in the system. Along with the gold standard, seven additional animations were created, each one missing or substituting a single illumination component.

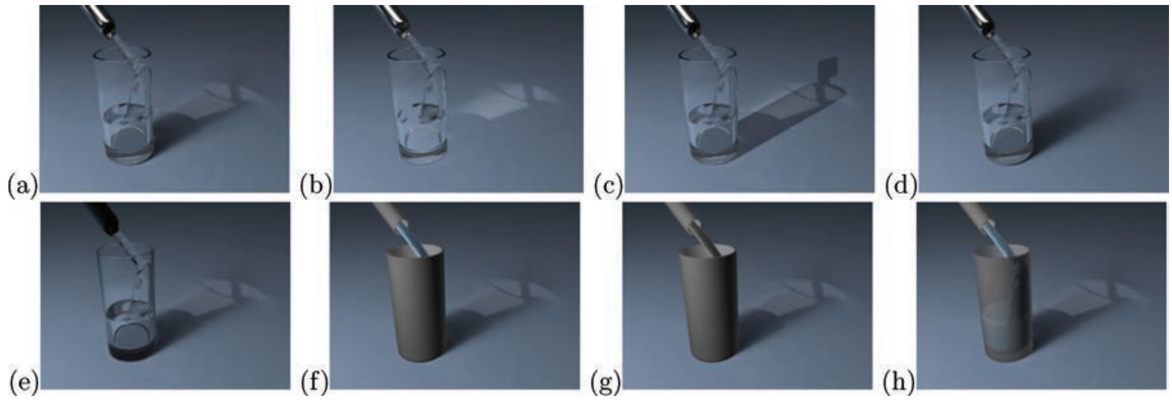

Fig. 2. One scene of the five used in our experiments. Figures (a) through (h) display different lighting effects as evaluated by one hundred participants. (a) Gold-Standard: includes all natural lighting effects; (b) No-Shadow: has no shadows; (c) No-SoftShadow: replaces soft shadows with hard shadows; (d) No-Caustics: has soft shadows but no caustics; (e) No-Specular: has no specular reflections; (f) Default-Color: includes only diffuse color instead of refraction; (g) Default-Grey: has no colors at all; and (h) Default-Transparency: shows default transparency.

The parameters tested included soft-shadow, hard-shadow, and no-shadow options for shadowing due to direct lighting. We used area light sources for soft shadows and point lights for hard shadows. Indirect lighting was included in each of the scenes with a virtual point light method to give consistency across the scenes. Lighting phenomena, such as caustics and specular highlights, were removed individually from the scene, and the refraction material properties of the water's surface was replaced with diffuse transparency, grey and color for faster rendering and simpler material properties.

These animations were then used to determine if visual differences exist with the absence or presence of each individual phenomenon. If a scene showed high visual impact with the removal or replacement of any of the parameters, these components were considered more important to the visual quality of the scene and contributed more to the visual accuracy of water. The animation titles are as follows: Gold-Standard, No-Shadow, No-SoftShadow, No-Caustics, No-Specular, Default-Color, Default-Transparency, and Default-Grey. Figure 2 shows one frame of the Glass Pour animation and labels for all of the animations.

3.2 Testing Scenes

Five testing scenes in the study were selected to give a wide range of various types of water dynamics and lighting conditions. Each scene contained all lighting elements and was rendered eight times to create a consistent basis for comparison. Moreover, every scene contained a diffuse object to convey refraction and reflection of light, and the surface collected caustics resulting from the water. The scenes are described below, and they are also on the accompanying video.

Glass Pour (Figure 2) is a common, discernible water type, with which every participant would have experience. This scene easily conveys caustics and is equipped with two strong light sources to display hard and soft shadows. It is a shallow-water scene with a low-water dynamic that does not present a complicated surface. There were four main materials present in the scene: a chrome material on the spout, the diffuse ground material, and two refractive materials; the water's index of refraction was set to 1.33 and 1.5 for the glass. Both the Glass Pour and the Dynamic Box scenes varied in their triangular polygon count, so the Glass Pour scene ranged from 151,000 to 246,000 triangles.

Dynamic Box (Figure 3(a)) is a similar shallow-water scene like Glass Pour, but it offers a higher range of dynamic water and contains a more complex water surface. Two light sources were used.

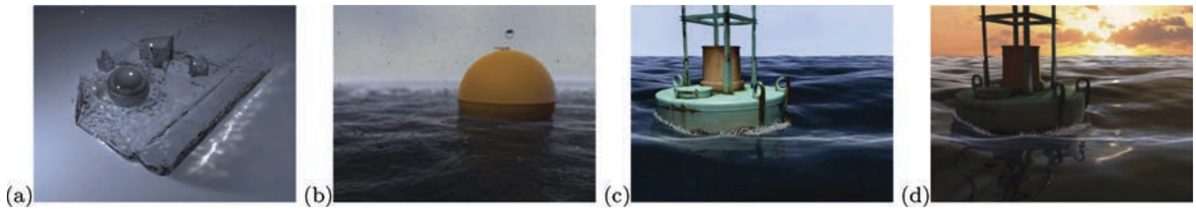

Fig. 3. Images of the testing scenes. The scenes were selected to show a wide variety of dynamic water, lighting, and material parameters. (a) The Dynamic Box; (b) Rainy Lake; (c) Sunny Buoy; and (d) Sunset Buoy.

Moreover, objects were placed in the water to show water dynamics and to depict hard and soft shadows. These objects were highly reflective to add additional emphasis to caustics. This scene varied from 99,000 to 434,000 triangles.

The *Rainy Lake* (Figure 3(b)) scene is a deepwater scene with a completely diffuse lighting setup. Diffuse lighting defocuses light rays and scatters them throughout the environment. An overcast sky gives less attention to the lighting phenomena and makes certain lighting phenomena more difficult to notice, such as shadows, specular highlights, and caustics. The water dynamic is low. The scene comprises 256,000 triangles and five different materials, with one being the refractive water material.

Sunny Buoy and Sunset Buoy (Figure 3(c) and (d)) were created as common, discernible deepwater scenes with only the lighting levels and positions altered. Sunny Buoy had a high dynamic range of lighting caused by the sun's direct illumination. It also generates harder shadows. The second scene contained a light source directly in front of the camera that conveyed better specular reflections and different caustics on the buoy. The buoy scenes were the most complex with regard to materials. The buoy was hand-painted with 24 materials dedicated to it alone, with textures ranging from 64 by 64 pixels to 1024 by 1024 pixels. The ocean materials contained both refractive and reflective quantities, which varied in value, to gain the appropriate stylistic appearance. Each scene had 955,000 triangles.

4. EXPERIMENTS

This section defines our testing metrics, and describes the testing system, population, environment, and other conditions of the study. We measure (1) the psychophysical test that shows the perceived difference in ranking of scene quality to determine which of the illumination components used in water animations are important for rendering; and (2) the rendering costs (time) associated with these illumination components using our rendering system and settings.

4.1 Procedure, System, Environment, and Population

We have developed a computer-based system for testing the material and lighting variables. It allows for eight animations to be simultaneously displayed and ranked, comments to be written, and results to be anonymously submitted. We have also carefully measured the rendering times for the cost analysis.

The testing steps were as follows. The participants were first given a brief demographic survey asking their experience level, age, gender, and presence of visual deficits, such as glasses or cataracts. After completing the survey, they were guided to each of the five scenes in randomized order. The eight animations of each scene were displayed randomly in a 4×2 pattern (see Figure 4). The animations were displayed in a loop for 75 seconds. This limited allotment was selected to uphold the criteria of the psychophysical analysis and to standardize the duration of viewing. A pilot study was conducted to verify that the assigned time was sufficient for the participants to make a decision. The exact question the subjects were asked was to "sort the animations according to their perceived visual quality," and they did this by ranking videos played on the screen from one to eight. The system constrained the

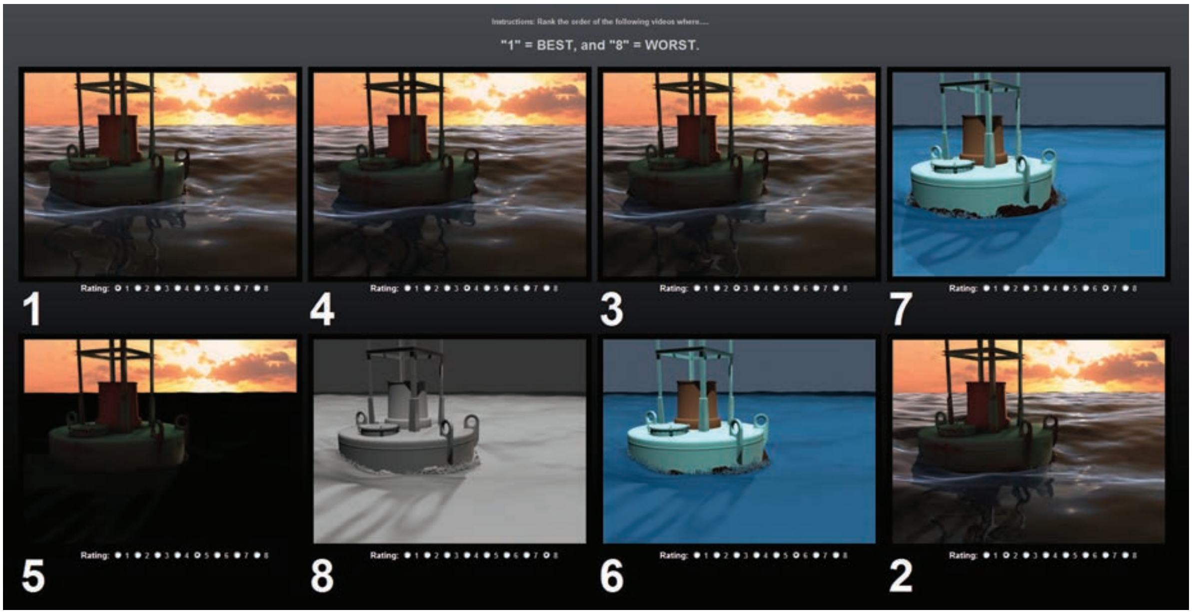

Fig. 4. The testing system displayed all animations of a single scene in randomized order; the participants ranked them according to the perceived visual quality. Their current choice is highlighted by a large number.

ranking to ensure that no video could be rated twice and that no two videos could be rated with the same value. After 75 seconds, the ranking was disabled and qualitative data was collected in text boxes beneath the videos. These boxes were used to collect any additional data the subjects felt would better clarify the decisions they had made.

The testing system used a 30″ monitor with a 2560 × 1600 resolution in which each video was displayed in a 480 × 360 resolution. This was the actual rendered size, and no scaling of the videos was performed. The screen resolution ensures that each animation was viewed with a sufficient detail and could be compared without loss of precision. The SMPTE standard 196M-2003 controlled the lab environment where testing was conducted [SMPTE 2003]. This standard provided requirements on viewing angle and projector luminance. Every participant conducted testing within the same environmental lighting and on the same computer system. Although the image size could appear small, our use of a 30″ monitor along with having participants seated close to the monitor made up for any possible lack of quality. In fact, participants remarked on the outstanding quality of the videos. It was very important for our testing parameters to display the videos adjacent to one another in a way that instantly revealed differences in lighting. We were considering an option of using multiple monitors or projectors, but we could not ensure a stable picture throughout the entire testing period, which spanned about three weeks. The final videos used a small compression to keep the file size small and thus enable smooth loading of the animation. We carefully selected a minimal video compression, because optimal video quality was of the utmost importance.

The study conducted testing on one hundred participants in the areas of science, engineering, and art. Introductory demographic information from the study revealed experience trends in our sample population. Testing was conducted on a university campus, and the participants were students, teachers, and staff personnel; some held levels of proficiency in the area of rendering, color, and lighting. Of the one hundred volunteers, 71% had some form of higher education. The majority of the participants

(94%) had no visual deficits, and no debilitating visual deficits were reported. Approximately half of the participants claimed they had at least a basic knowledge of light theory and the rendering process.

4.2 Psychophysical Analysis

A psychophysical analysis (PPA) is a comparative study used to discern perceivable differences among stimuli. Instead of a traditional scalar quantification of lone stimuli as in previous studies—which humans are extremely poor at judging—the psychophysical analysis allows for an ordinal pairing between stimuli or values, and judgments are made on the contrast between them. In our study we use a PPA to measure the perceived visual quality of the lighting components and material variables in animations. Each animation of a scene was ranked on a scale of one to eight, one being the highest visual quality. Subjects were able to view all stimuli at once instead of making judgments on pairs of stimuli only. This also drastically cut the time needed to complete the study and subsequently decreased the number of tested subjects. Raw data generated by the psychophysical analysis was of a discrete ordinal nature, which offered no continuum to the quantitative data.

Other than the gold standard, all animations are defined by the material or lighting component that they lack. By statistically analyzing the difference between pairs of animation average rankings, we establish a relationship of importance between pairs of lighting phenomena. We will interchangeably use the words animation and lighting phenomenon.

4.3 Rendering Cost

The actual rendering cost of each phenomenon was determined by an in-depth analysis of the Mental Ray rendering logs. To diminish the influence of the operating system, hard-drive swaps, memory, and CPU cache misses, and so on, we rendered 30 consecutive frames from the middle of the animation for each of the lighting components and calculated averages of the rendering times. This also helps to better exploit the scene time coherence, and it shows better performance of the renderer than if only individual frames are rendered. The relative rendering time was calculated by comparing the individual results to another animation that had only one lighting phenomenon changed. This revealed the individual component cost and a relative percentage cost versus all lighting phenomena. These measurements were conducted for each of the five testing scenes a total of three times for each phenomenon. A total of 3,600 frames were rendered.

While rendering the final compilation for each scene, it was chosen to render each lighting phenomena to a different rendering layer. Doing this allowed the researchers to isolate rendering performance contributing to each lighting effect and to easily compile the various animations. All scenes were rendered with the same basic lighting setup, though each scene required a different number of lights and settings for effects such as caustic photons. Lighting was created with a series of spotlight sources to simulate light bounce throughout the scene without incurring the additional cost of other bounce approximations like final gather. Hard shadows were created using depth map shadows, and soft shadows used ray-tracing and area lights. Caustics employed photon mapping to render accumulation highlights due to the bouncing and refraction of light. There were various amounts of photons used for each scene and the researchers added photons up to the point of diminishing returns. The goal was to render the effect accurately without incorrectly recording the rendering cost. The rest of the effects were controlled through the water's shader. Default grey, and color used a diffuse shader; additionally transparency, changed the opacity channel on the color shader. Specular highlighting used the water shader, but altered the shader's specular channel. During rendering we used raytracing and photon maps in Mental Ray. We used production settings for final trace that included antialiasing, raytracing with seven reflection and refraction bounces, BSP2 for intersection evaluation acceleration, and we also used HDR for image-based lighting.

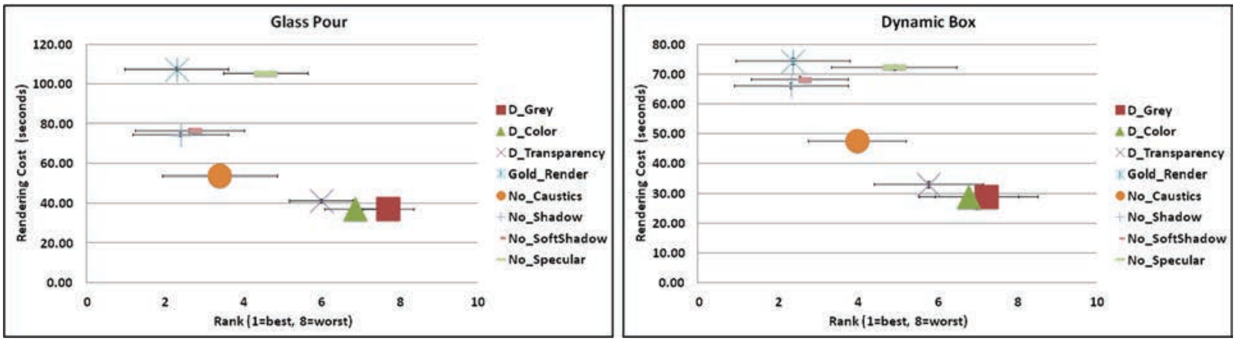

Fig. 5. Average ranking versus rendering cost: Glass Pour (left) and Dynamic Box.

5. RESULTS AND FINDINGS

In this section we describe results of our psychophysical analysis. All graphs in this section show on the horizontal axis the average ranking of the animations for the specified scene, and on the vertical axis the average rendering cost of the animations. The standard deviation of the rankings is shown in the error bars accompanying the plot. By analyzing the graphs, we determine cost-versus-benefit and recommend savings by means of reducing the rendering time.

For example, rendering the Rainy Lake scene (Figure 9) without caustics, takes 41 seconds compared to the Gold-Standard with 72 seconds of rendering time, but they have similar rankings. The statistical test described below confirms that their rankings show no significant difference, which suggests that there is not enough evidence to show that they are ranked differently. Similarly, rendering the Glass Pour scene (Figure 5) with no specular reflections would take 105 seconds compared to 107 seconds for the Gold-Standard but the statistical test shows significant difference between their rankings. We have performed two-sample ranked Wilcoxon tests [Walpole and Myers 1985] with a significance level of 0.05 where we compare visual difference among scenes. This test can be used to work with data that cannot be assumed to be normal such as our rankings. Our null hypothesis for the tests is that the mean ranking of one animation is equal to the mean ranking of a second animation. We applied this null to each pair of animations; this is done for every scene independently. In the rest of this article, whenever the rankings of two animations are said to have no significant difference, it can be interpreted as showing no sufficient statistical evidence to reject the null hypothesis. The results of the Wilcoxon tests comparing the Gold-Standard, No-Caustics, No-Shadow, No-SoftShadow and No-Specular are reported in Tables I to V. The rest of the results are referenced when appropriate, since in most cases their mean rankings were very low.

5.1 Psychophysical Analysis

5.1.1 *Glass Pour and Dynamic Box.* Those two commonly recognizable scenes study the differences between low- and high-water dynamics in nearly identical lighting conditions. These results show (see Figure 5) an important trend in the data seen across all animations. As mentioned above, a statistical test was conducted between the ratings of each animation to compare their average ranking. Most participants ranked the Gold-Standard and No-Shadow as having no significant statistical difference in their average ranking for Glass Pour and Dynamic Box, as can be seen in Table I. This translates into an average 74 seconds of rendering time for No-Shadow as opposed to 107 seconds for the Gold-Standard in the case of the Glass Pour scene and 66 versus 74 seconds respectively for the Dynamic Box scene.

Table I. Mean, Standard Deviation, and Two-Sample Ranked Wilcoxon Tests for the Glass Pour Scene

GLASS POUR	Gold-Standard	No-Caustics	No-Shadow	No-Soft Shadow	No-Specular
Mean	2.31	3.40	2.41	2.64	4.57
Std dev.	1.32	1.47	1.21	1.38	1.07
Wilcoxon test					
No-Caustics	$Z = -4.38$ $p < 0.0001$				
No-Shadow	$Z = -0.06$ $\mathbf{p = 0.47}$	$Z = 4.47$ $p < 0.0001$			
No-SoftShadow	$Z = -1.34$ $\mathbf{p = 0.08}$	$Z = 3.45$ $p = 0.0002$	$Z = -0.87$ $\mathbf{p = 0.19}$		
No-Specular	$Z = -7.57$ $p < 0.0001$	$Z = -4.57$ $p < 0.0001$	$Z = -7.79$ $p < 0.0001$	$Z = -7.29$ $p < 0.0001$	

Table II. Mean, Standard Deviation and Two-Sample Ranked Wilcoxon Tests for the Dynamic Box Scene

DYNAMIC BOX	Gold-Standard	No-Caustics	No-Shadow	No-SoftShadow	No-Specular
Mean	2.38	3.98	2.33	2.54	4.91
Std dev.	1.43	1.22	1.42	1.22	1.57
Wilcoxon test					
No-Caustics	$Z = -6.27$ $p < 0.0001$				
No-Shadow	$Z = 0.43$ $\mathbf{p = 0.33}$	$Z = 6.00$ $p < 0.0001$			
No-SoftShadow	$Z = -1.17$ $\mathbf{p = 0.12}$	$Z = 6.35$ $p < 0.0001$	$Z = -1.31$ $\mathbf{p = 0.09}$		
No-Specular	$Z = -7.78$ $p < 0.0001$	$Z = -4.05$ $p < 0.0001$	$Z = -7.19$ $p < 0.0001$	$Z = -7.47$ $p < 0.0001$	

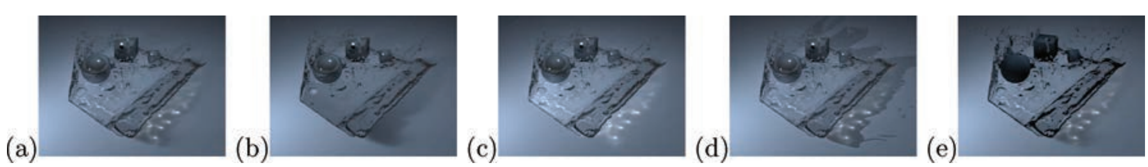

Fig. 6. Dynamic Box scene. (a) Gold-Standard; (b) No-Caustics; (c) No-Shadow; (d) No-SoftShadow; (e) No-Specular.

These shallow water scenes showed a higher visual quality associated with caustics and specular reflections than all shadowing components. Caustics ranked higher in the shallow-water scenes and were more easily perceived when absent as seen in Table I and Table II. With the higher dynamics of water, the Dynamic Box scene showed an increased visual importance of caustics over the Glass Pour; No-Caustics Dynamic Box (M=3.98 SD=1.22), No-Caustics Glass Pour (M=3.40 SD=1.47); $Z = -3.36$, $p=0.0003$. This is attributed to the increased velocity and to a more complex water surface.

All animations not rendered with refraction were ranked as the lowest visual quality. These animations (i.e., Default-Grey, Default-Color, and Default-Transparency) are best for previsualizations, and as much as 50% of the rendering time can be saved. Figure 6 shows the most important effects for the Dynamic Box scene and Figure 2 shows all the effects for the Glass Pour scene.

Table III. Mean, Standard Deviation and Two-Sample Ranked Wilcoxon Tests for the Sunny Buoy Scene

SUNNY BUOY	Gold-Standard	No-Caustics	No-Shadow	No-SoftShadow	No-Specular
Mean	2.60	2.71	2.40	2.52	6.66
Std dev.	1.15	1.30	1.24	1.22	1.35
Wilcoxon test					
No-Caustics	$Z = -0.53$ **$p = 0.29$**				
No-Shadow	$Z = 1.02$ **$p = 0.15$**	$Z = 1.33$ **$p = 0.09$**			
No-SoftShadow	$Z = 0.13$ **$p = 0.444$**	$Z = 0.67$ **$p = 0.25$**	$Z = -0.76$ **$p = 0.22$**		
No-Specular	$Z = -8.25$ $p < 0.0001$	$Z = -8.47$ $p < 0.0001$	$Z = -8.49$ $p < 0.001$	$Z = -8.72$ $p < 0.001$	

Table IV. Mean, Standard Deviation and Two-Sample Ranked Wilcoxon Tests for the Sunset Buoy Scene

SUNSET BUOY	Gold-Standard	No-Caustics	No-Shadow	No-SoftShadow	No-Specular
Mean	2.57	2.93	2.23	2.49	6.06
Std dev.	1.10	1.15	1.47	1.05	1.32
Wilcoxon test					
No-Caustics	$Z = -2.04$ $p = 0.02$				
No-Shadow	$Z = 1.46$ **$p = 0.071$**	$Z = 3.34$ $p = 0.0004$			
No-SoftShadow	$Z = 0.79$ **$p = 0.21$**	$Z = 2.32$ $p = 0.01$	$Z = -1.51$ **$p = 0.06$**		
No-Specular	$Z = -8.50$ $p < 0.0001$	$Z = -8.51$ $p < 0.0001$	$Z = -8.42$ $p < 0.001$	$Z = -8.59$ $p < 0.001$	

5.1.2 *Sunny and Sunset Buoy.* These two scenes shared identical water dynamics and contain a recognizable, high dynamic range of lighting conditions. As in the previous scene, the Gold-Standard also showed no significant statistical difference with No-Shadow (Table III and Table IV). No significant differences were found between the Gold-Standard and the animation with hard shadows. They showed little variation among the four most visually accurate animations. This was also reiterated in the qualitative data collected in the text fields where participants expressed their inability to distinguish differences between these animations. In the scene Sunny Buoy, the Gold-Standard ranking showed no significant statistical difference against the animation without caustics. Also in this scene, there was no significant difference between the animations with no shadows, no hard shadows and no caustics, as seen in Table III. With limited diffuse surfaces in the scenes, variations in shadow and caustics were difficult to distinguish. This translates into a savings of 20% to 40% of rendering time compared to the time taken to render the Gold-Standard animation.

In contrast to shadows and caustics, specular reflection showed high importance in visual accuracy of the deepwater scenes. For Sunny Buoy, No-Specular was rated as low as the videos without refraction, which otherwise was consistently the most important illumination component of the scenes. In the Sunny Buoy scene, for example, the animation with no specular reflection ($M = 6.66$ $SD = 1.35$) was rated as low as the one with default grey ($M = 6.93$ $SD = 1.24$); $Z = 1.2$, $p = 0.09$. Note that Default-Grey does not appear in the tables. Specular reflection has both high visual importance and

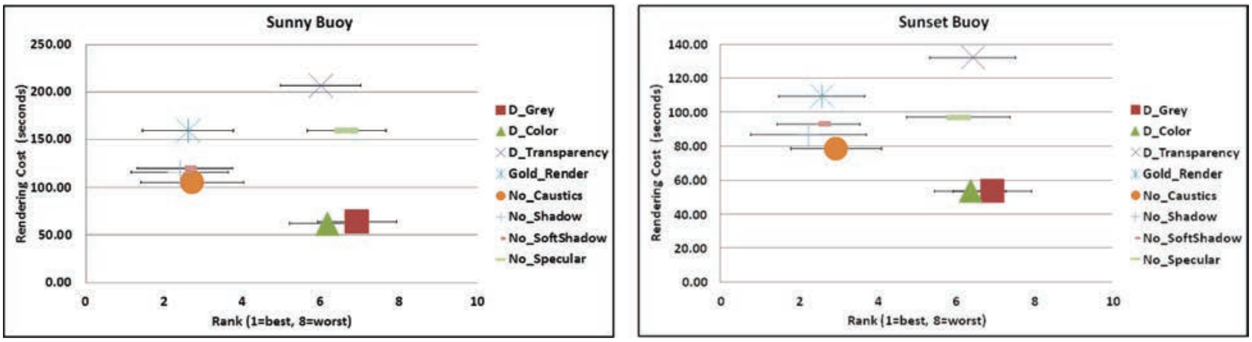

Fig. 7. Average ranking versus rendering cost: Sunny Buoy (left) and Sunset Buoy.

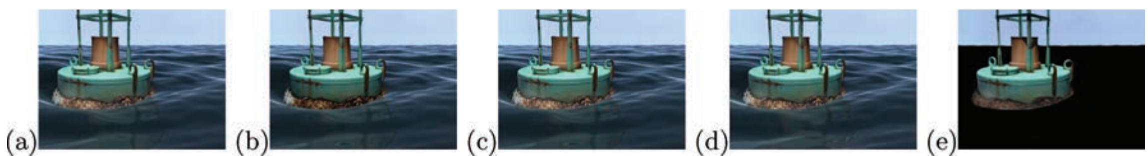

Fig. 8. Sunny Buoy scene. (a) Gold-Standard; (b) No-Caustics; (c) No-Shadow; (d) No-SoftShadow; (e) No-Specular.

Table V. Mean, Standard Deviation and Two-Sample Ranked Wilcoxon Tests for the Rainy Lake Scene

RAINY LAKE	Gold-Standard	No-Caustics	No-Shadow	No-SoftShadow	No-Specular
Mean	2.51	2.86	2.54	2.54	6.88
Std dev.	1.29	1.52	1.17	1.17	1.29
Wilcoxon test					
No-Caustics	$Z = -1.54$ **$p = 0.06$**				
No-Shadow	$Z = -0.076$ **$p = 0.35$**	$Z = 1.06$ **$p = 0.14$**			
No-SoftShadow	$Z = -0.38$ **$p = 0.35$**	$Z = 1.30$ **$p = 0.14$**	$Z = 0.16$ **$p = 0.43$**		
No-Specular	$Z = -8.54$ $p < 0.0001$	$Z = -8.44$ $p < 0.0001$	$Z = -8.36$ $p < 0.001$	$Z = -8.44$ $p < 0.001$	

low rendering cost, so for all scenes, deepwater scenes in particular, specular reflection is very important and should be used. Figure 8 shows the most important effects for the Sunny Buoy scene and Figure 4 shows all the effects for the Glass Pour scene.

5.1.3 *Rainy Lake.* This scene studies the effect of diffuse lighting and low water dynamics. In these scene ratings the Gold-Standard showed no significant difference to the animations with no caustics and no shadows (see Table V). For this scene, 31% participants expressed the inability to perceive differences. Savings coming from omitting shadows or caustics would conceivably result in up to 43% savings with respect to the Gold-Standard rendering time. As with the Sunny Buoy scene No-Specular ($M = 6.88$ $SD = 1.29$) was rated as the least visually pleasing animation with Default-Grey ($M = 6.82$ $SD = 1.27$); $Z = -0.01$, $p = 0.49$, which consistently resulted in the lowest ranking throughout the rest of the scenes. Similar low rankings for No-Specular in the other deepwater scenes emphasizes its high visual importance in scenes of this nature. This illumination component is a great candidate for

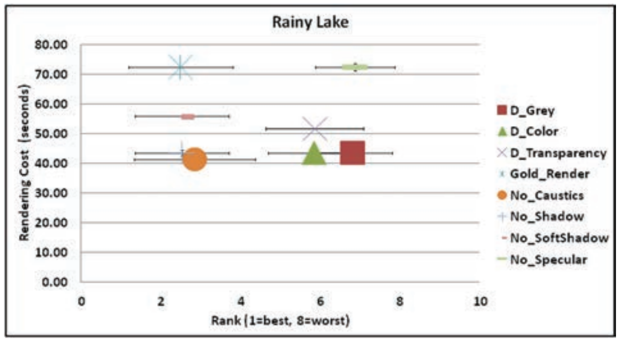

Fig. 9. Average ranking versus rendering cost of the scene Rainy Lake.

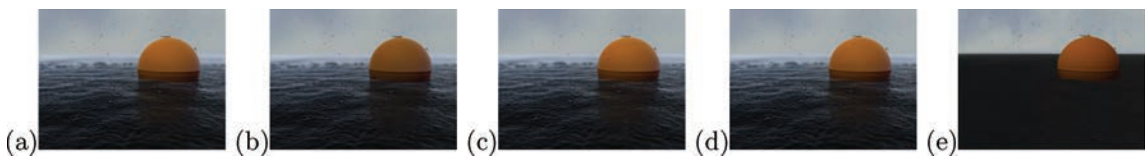

Fig. 10. Rainy Lake scene. (a) Gold-Standard; (b) No-Caustics; (c) No-Shadow; (d) No-SoftShadow; (e) No-Specular.

increasing visual quality with low effect on rendering time. Figure 10 shows the most important effects for the Rainy Lake scene.

5.2 Video Quality Metric

In order to provide further comparison we used the online tool Dynamic Range Independent Video Quality Assessment and Dynamic Range Independent Metric Online (DRIVQM) which measures visual differences between two (sequences of) images. Significantly different pixels are displayed in red; identical pixels are unchanged; and pixels that are close to just-noticeable difference are displayed using a predefined colormap [Aydin et al. 2010].

Figure 11(a) shows results of the comparison of five different frames from the scene Gold-Standard SunnyBuoy against the same scene rendered without caustics. Similarly, Figure 11(b) compares the Gold-Standard against the scene without shadows. Figure 11(c) shows results of comparison of five frames from the scene Gold-Standard Dynamic Box compared to the same sequence without soft shadows. There are areas of significantly different colors (displayed in red), but most of the pixels are either identical or slightly different as can be seen in the case of caustics in Figure 11(a) or missing shadows in Figure 11(c). These results show comparisons of two sequences and have only an informative character, as further analysis is described below. The overall variance suggests there are slight differences in areas that can be predicted such as missing caustics, shadows, and so on.

To quantify the color-coded video sequences returned by DRIVQM, the percentage of significantly different pixels in each frame was calculated. The significantly different pixels are those for which the red component of RGB is higher than 50%, and the green and blue components are less than 50%. Table VI shows the percentage of red pixels in the video sequences. As can be seen, the Glass Pour scene shows as much as 35% of red pixel coverage as well as the Dynamic Box scene. Sunny and Sunset Buoy present less than 5% difference, in any of the considered lighting conditions. Rainy Lake shows a very small number of red pixels when no caustics are present and 19% of red pixel coverage when no shadows or no soft shadows are applied.

Table VI. Red Pixel Percentage in DRIVQM Comparison

	No-Caustics	No-Shadow	No-SoftShadow
Glass Pour	35%	29%	32%
Dynamic Box	25%	34%	16%
Sunny Buoy	4%	3%	1%
Sunset Buoy	2%	4%	0.4%
Rainy Lake	3%	19%	19%

Table VII. Rendering Cost in Seconds of the Five Scenes and Their Animations

Scene/Anim.	Def.-Col.	Def.-Grey	Def.-Transp.	Gold-Std.	No-Caus.	No-Shad.	No-SoftShad.	No-Spec.
Glass Pour	37.24	37.24	41.38	107.58	53.79	74.48	76.55	105.51
Dynamic Box	28.97	28.97	33.10	74.48	47.59	66.21	68.28	72.41
Sunny Buoy	62.07	64.14	206.90	159.31	105.52	115.86	120.00	159.31
Sunset Buoy	53.76	53.76	132.38	109.65	78.62	86.90	93.10	97.27
Rainy Lake	43.45	43.45	51.73	72.41	41.38	43.45	55.86	72.41

It is interesting to see that while the psychophysical results were generally consistent across all the scenes, the VQM measure varied between types of scenes. The shallow water scenes present a much higher red pixel percentage compared to the deepwater scenes (Table VI). We attribute this to the pixel to pixel comparison of the DRIVQM test. The shallow water scenes present considerable differences when shadows or caustics are not present (Figures 2 and 6). In contrast, the deepwater scenes present a more subtle difference, given that there are not many objects casting shadows or reflecting caustics (for example, see Figures 8 and 10).

The percentage of red pixel coverage for Glass Pour and Dynamic Box suggests that there is a difference between the animations of each of those scenes. However, the psychophysical analysis suggests that the test subjects found no significant difference between some of those animations. Finally, the Sunny and Sunset Buoy animations with No-Caustics, No-Shadow, and No-SoftShadow depict small differences with the Gold-Standard, which is also suggested by the psychophysical results in which even the render without caustics has rankings that show no significant difference with the Gold-Standard rankings.

In the psychophysical analysis, the Rainy Lake scene showed no significant difference between the Gold-Standard and No-Caustics, No-Shadow and No-SoftShadow. The DRIVQM test suggests a different behavior (see Table VI). We conclude from these results that while the DRIVQM measure showed a large difference between shallow water and deepwater scenes, it was not possible to tell if there exists a significant difference within animations of the same scene and their Gold-Standard, given that the results from the DRIVQM measure were not always consistent with the psychophysical test.

The DRIVQM metric is used to measure a just-noticeable difference, and: compares significantly similar and different pixels by their values. It does not work for some subtle effects in moving water, and it cannot capture long-range influences, such as long-cast shadows or caustics.

5.3 Rendering Cost

The rendering cost is displayed in all graphs and is also summarized in Table VII. Each number in the table represents the time in seconds of the rendering of the given scene. The highest savings are achieved if the scene is rendered without refraction, but caustics and shadows also have a high rendering cost.

Total rendering costs of each scene varied with the amount and types of lighting, the scene polygon count, and complexity. The Dynamic Box scene was the least complex and fastest to render at an

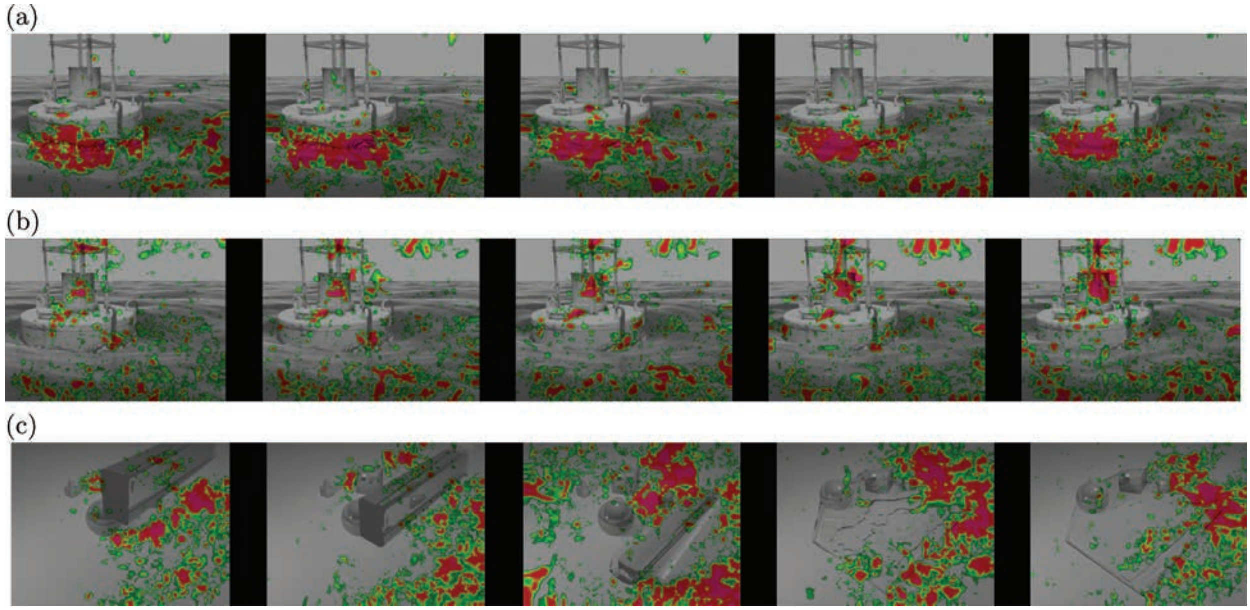

Fig. 11. DRIVQM comparison. (a) Sunny Buoy Gold-Standard and No-Caustics; (b) Sunny Buoy Gold-Standard and No-Shadow; (c) Dynamic Box Gold-Standard and No-SoftShadows.

average of 76.55 seconds per frame. The rest of the scenes in order were 84.83 seconds for the Rainy Lake scene; 109.65 for the Glass Pour Scene; 115.86 for the Sunset Buoy Scene; and 163.45 for the Sunny Buoy Scene. A total average cost of 110.07 seconds per frame was calculated for all scenes. By combining the rendering cost and the results of the psychophysical analysis, we can conclude that by not rendering shadows, we can save an average of 21%–26% of the rendering time compared to rendering the Gold-Standard animation. For some scenes, such as the Glass Pour, these savings can even be around 30%. A commonly used previsualization approach, rendering a scene using only diffuse materials, will sacrifice a significant part of visual quality, but will save only about 30%–50% of time compared to the Gold-Standard animation.

6. CONCLUSIONS AND FUTURE WORK

We have developed a system for testing and comparing the visual importance of lighting phenomena in animated sequences of water in the most typical environments, ranging from low to high dynamics of water motion and illumination. We have tested five different scenes, each with eight variations of illumination components. One hundred participants evaluated the scenes by ranking them on a scale of best-to-worst according to the perceived visual quality. We have attempted to provide the best possible testing conditions by using a strictly controlled testing environment. The testing conditions were equal for all participants in terms of timing and delivery of results.

Our results suggest that for scenes similar to the ones tested, no shadow and no soft shadows were indistinguishable from the Gold-Standard. This is concluded from the results of statistical testing which found no statistically significant difference between the Gold-Standard animation and the animations with no shadows in every scene tested. The traditional scenes used in previsualization, such as rendering the water material property as a grey or colored diffuse surface, are generally perceived as nonrealistic, but can lead to savings in rendering time of up to 50% compared to the time required to render the Gold-Standard animation. On the other hand, the most important lighting phenomenon related to water according to our participants is reflection.

We have found that for our scenarios, rendering system, settings, and hardware, it is possible to save up to an average of 20% of rendering time by not rendering the effects that are not considered important. The actual perceived visual difference to the Gold-Standard is negligible according to the statistical tests. For scenes with low-water dynamics and high dynamic range, the caustics and shadows can be dropped, resulting in nearly 40% additional savings in rendering time compared to the render time for the Gold-Standard with almost no sacrifice of the visual quality. The speedup is relative to the speed of the computer being used. We believe that perceptually driven rendering can bring about an important improvement in rendering times and perceived visual quality for computer animations. Large rendering times can be reduced by careful psychophysical analysis and by considering only the most relevant phenomena. However, our work can be improved in many ways.

We have attempted to choose a range of scenes that is as wide as possible, and we believe that the consistency of our results over widely different scenes is a strong argument for our findings. However, many different scenes can be chosen, and there is always a chance that they could possibly show different results. Other possible future work is in the direction of mutual dependency of the evaluated phenomena. Because of the requirement for many participants, we were unable to analyze the effects of multiple missing or substituted phenomena. It is highly probable that such second-order dependencies exist, but they are impossible to detect with our experiments. Many more animations of the same scene would be necessary to create such an experiment, and participants in the study would possibly be overwhelmed by the choices. Our experiment requires sorting the sequences. A better approach, although requiring more time to perform the experiment, would involve pairwise comparison [Ledda et al. 2005]. Another avenue for future work would be to include more complex water phenomena, such as bubbles and foam. Interesting discoveries could be achieved by tracking the human eye and considering only the salient features of the animations. It would also be important to see how subsurface scattering and participating media influence the perceived visual quality. Also, it would be interesting to study how the gender and age as well as previous knowledge of computer graphics affect the perceived quality of the rendering.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments and suggestions that significantly helped to improve the manuscript. We would also like to thank Gary Bertoline for his comments on the used technique, Kellen Meicher for help in the design of the system, and Luis Barboza for help with the statistical analysis.

REFERENCES

ANSON, O., SUNDSTEDT, V., GUTIERREZ, D., AND CHALMERS, A. 2006. Efficient selective rendering of participating media. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (APGV '06), ACM, New York, 135–142.

AYDIN, T. O., ČADÍK, M., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Video quality assessment for computer graphics applications. In *ACM SIGGRAPH Asia 2010 Papers*. (SIGGRAPH ASIA '10) ACM, New York, 161:1–161:12.

CATER, K., CHALMERS, A., AND DALTON, C. 2003. Varying rendering fidelity by exploiting human change blindness. In *Proceedings of the 1st International Conference on Computer Graphics and interactive Techniques in Australasia and South East Asia (GRAPHITE '03)*, ACM, New York, 39–46.

CATER, K., CHALMERS, A., AND LEDDA, P. 2002. Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '02)*, ACM, New York, 17–24.

CHALMERS, A., DEBATTISTA, K., AND DOS SANTOS, L.P. 2006. Selective rendering: Computing only what you see. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE '06)*, ACM, New York, 9–18.

DEBATTISTA, K., SANTOS, L., AND CHALMERS, A. 2006. Accelerating the irradiance cache through parallel component-based rendering. In *The 6th Eurographics Symposium on Parallel Graphics Visualization* (EGPGV '06), 27–34.

DEBATTISTA, K., SUNDSTEDT, V., SANTOS, L. P., AND CHALMERS, A. 2005. Selective component-based rendering. In *Proceedings of the 3rd International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE'05)*, ACM, New York, 13–22.

DIDYK, P., EISEMANN, E., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Perceptually-motivated real-time temporal upsampling of 3D content for high-refresh-rate displays. In *Comput. Graph. Forum* (Eurographics 2010).

DMITRIEV, K., BRABEC, S., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2002. Interactive global illumination using selective photon tracing. In *Proceedings of the 13th Eurographics Workshop on Rendering (EGRW '02)*, Eurographics Association, 25–36.

DONNER, C. AND JENSEN, H. W. 2005. Light diffusion in multi-layered translucent materials. *ACM Trans. Graph. 24*, 3, 1032–1039.

DONNER, C., LAWRENCE, J., RAMAMOORTHI, R., HACHISUKA, T., JENSEN, H. W., AND NAYAR, S. 2009. An empirical bssrdf model. *ACM Trans. Graph. 28*, 3, 1–10.

DUMONT, R., PELLACINI, F., AND FERWERDA, J. A. 2003. Perceptually-driven decision theory for interactive realistic rendering. *ACM Trans. Graph. 22*, 2, 152–181.

EL-NASR, M. S. AND YAN, S. 2006. Visual attention in 3D video games. In *Proceedings of the ACM SIGCHII international Conference on Advances in Computer Entertainment Technology (ACE '06)*. ACM, New York, 22.

ENRIGHT, D., MARSCHNER, S., AND FEDKIW, R. 2002. Animation and rendering of complex water surfaces. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '02)*. ACM, New York, 736–744.

HACHISUKA, T. AND JENSEN, H. W. 2009. Stochastic progressive photon mapping. In *ACM SIGGRAPH Asia Papers*. ACM, New York, 1–8.

HACHISUKA, T., OGAKI, S., AND JENSEN, H. W. 2008. Progressive photon mapping. In *ACM SIGGRAPH Asia Papers*. ACM, New York, 1–8.

HASIC, J. AND CHALMERS, A. 2009. Saliency in motion-picture. In *Proceedings of the Spring Conference on Computer Graphics (SCCG'09)*. 186–195.

HASIC, J., CHALMERS, A., DEBATTISTA, K., AND MASTROROPOLOU, G. 2007. Movement bias in visual attention for perceptually guided selective rendering in animations. In *Proceedings of the Spring Conference on Computer Graphics (SCCG'07)*.

IRAWAN, P., FERWERDA, J. A., AND MARSCHNER, S. R. 2005. Perceptually based tone mapping of high dynamic range image streams. In *Proceedings of the 16th Eurographics Workshop on Rendering Techniques*. 231–242.

JENSEN, H. W. 2001. *Realistic Image Synthesis Using Photon Mapping*. A. K. Peters, Ltd., Natick, MA.

KOZLOWSKI, O. AND KAUTZ, J. 2007. Is accurate occlusion of glossy reflections necessary? In *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization (APGV '07)*. ACM, New York, 91–98.

LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. In *ACM SIGGRAPH Papers*. ACM, New York, 640–648.

MASIA, B., AGUSTIN, S., FLEMING, R. W., SORKINE, O., AND GUTIERREZ, D. 2009. Evaluation of reverse tone mapping through varying exposure conditions. In *ACM SIGGRAPH Asia Papers*. ACM, New York, 1–8.

MCDONNELL, R., LARKIN, M., DOBBYN, S., COLLINS, S., AND O'SULLIVAN, C. 2008. Clone attack! Perception of crowd variety. In *ACM SIGGRAPH Papers*. ACM, New York, 1–8.

MCDONNELL, R., LARKIN, M., HERNÁNDEZ, B., RUDOMIN, I., AND O'SULLIVAN, C. 2009. Eye-catching crowds: saliency based selective variation. *ACM Trans. Graph. 28*, 3, 1–10.

MCNAMARA, A. M. 2005. Exploring perceptual equivalence between real and simulated imagery. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization (APGV '05)*. ACM, New York, 123–128.

RAMANARAYANAN, G., BALA, K., AND FERWERDA, J. A. 2008. Perception of complex aggregates. *ACM Trans. Graph. 27*, 3, 1–10.

SATTLER, M., SARLETTE, R., MÜCKEN, T., AND KLEIN, R. 2005. Exploitation of human shadow perception for fast shadow rendering. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization (APGV '05)*. ACM, New York, 131–134.

SMPTE. 2003. SMPTE 196m-2003 *Motion-Picture Film Indoor Theater and Review Room Projection Screen Luminance and Viewing Conditions*. The Society for Motion Picture and Television Engineers.

STOKES, W. A., FERWERDA, J. A., WALTER, B., AND GREENBERG, D. P. 2004. Perceptual illumination components: a new approach to efficient, high quality global illumination rendering. In *ACM SIGGRAPH Papers*. ACM, New York, 742–749.

Tolhurst, D. J., Ripamonti, C., Párraga, C. A., Lovell, P. G., and Troscianko, T. 2005. A multiresolution color model for visual difference prediction. In *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization (APGV '05)*. ACM, New York, 135–138.

Walpole, R. E. and Myers, R. H. 1985. *Probability and Statistics for Engineers and Scientists* 3rd Ed. Macmillan.

Yu, I., Cox, A., Kim, M. H., Ritschel, T., Grosch, T., Dachsbacher, C., and Kautz, J. 2009. Perceptual influence of approximate visibility in indirect illumination. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization (APGV '09)*. ACM, New York.

Received July 2011; revised August 2012; accepted August 2012